# Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
*Department of Psychology, University of Michigan*
*Ann Arbor, Michigan*

DAVID A. GRANT, Associate Editor
*University of Wisconsin*

WILLIAM K. ESTES, Associate Editor
*Indiana University*

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

## CONTENTS

Types of Conflict and Their Resolution: A Reinterpretation: J. M. ATTHOWE, JR. ..... 1

Age Differences in Card-Sorting Performance in Relation to Task Difficulty, Task Set, and Practice: J. BOTWINICK, J. S. ROBBIN, AND J. F. BRINLEY ..... 10

Onset Versus Termination of a Stimulus as the CS in Eyelid Conditioning: W. J. HANSCHE AND D. A. GRANT 19

The Somatic Background of Rote Learning: R. N. BERRY AND R. C. DAVIS ..... 27

"Association" of Nonsense Syllables Following Varied Learning Conditions: B. W. CAMP .. 35

Predictions of Sequences of Successes and Failures: C. B. DESOTO, E. B. COLEMAN, AND P. L. PUTNAM 41

Gradients of Generalization in Secondary Reinforcement: B. O. BERGUM ..... 47

Retention of Responses to Stimulus Classes and to Specific Stimuli: K. E. LLOYD ..... 54

Growth of Sensation on Seven Continua as Measured by Force of Handgrip: J. C. STEVENS, J. D. MACK, AND S. S. STEVENS 60

The Relationship between Direction and Amount of Stimulus Change and Amount of Perceptual Disparity Response: H. D. KIMMEL ..... 68

American Psychological Association

Vol. 59, No. 1 January 1960

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1960 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

---

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1960 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 59, NO. 1 JANUARY, 1960

# TYPES OF CONFLICT AND THEIR RESOLUTION: A REINTERPRETATION [1]

JOHN M. ATTHOWE, JR.

*University of Alabama*

The effect of differing types of choice situations upon decision making was investigated. The subclass of choice situations examined was that of conflict. Conflict was defined as that situation of choice in which two alternatives are reacted to as approximately equal in preference. Types of conflict were defined in terms of the expected outcomes (anticipated reinforcements) associated with each alternative. Outcomes were designated as either attractive or repulsive, after Lewin (1935). In the present study, attraction or approach was equated with the possibility of gaining money. Repulsion or avoidance was equated with the possibility of losing money.

Lewin (1935) described three discrete types of conflict which he called the approach-avoidance (*ap-av*), *ap-ap*, and *av-av* types. A fourth, the double *ap-av* type, has been described by Hovland and Sears (1938). This latter type is an extension of Lewin's *ap-av* conflict. In the double *ap-av* type, *S* is influenced by the possible consequences of both an *ap* and an *av* outcome associated with each alternative. The weighing of four possible outcomes not only introduces a more complex set of decision operations, but also it is more nearly representative of the conflicts faced in everyday life.

The four types of conflict described by Lewin and Hovland and Sears have often been considered prototypes of conflict. The present paper views types of conflict not as discrete situations but as points along a dimension or continuum defined by the relative magnitudes of the *ap* and *av* outcomes. Position on this continuum is specified by the following ratio:

$$\frac{ap - av}{ap + av}$$

*Ap* and *av* refer to the total magnitudes of these tendencies within the situation of choice. This ratio is applicable only to situations of indifference (i.e., conflict). As the *ap*

[1] This article is based upon a dissertation submitted to the Department of Psychology, Stanford University in partial fulfillment of the requirements for the Ph.D. degree. The author is grateful to R. R. Sears, P. R. Farnsworth, L. Festinger, K. Little, and P. S. Siegel for their suggestions and criticisms.

outcomes come to dominate, this ratio approaches the limit of +1.00. As the *av* outcomes come to dominate, the ratio approaches the limit of −1.00. At one end of the continuum lies the *ap-ap* and multiple *ap-ap* conflicts; at the other pole lies the *av-av* and multiple *av-av* types. In other words, at +1.00 *S* is sure to gain; at −1.00 he is sure to lose.[2] At the midpoint, the relative magnitudes of the *ap* and *av* outcomes (i.e., gains vs. losses) are equal; here the ratio is equal to 0.00. This is the point of "complete ambivalence." Between the midpoint and the *ap* pole lies an array of conflict types, varying only in the relative magnitude of the *ap* and the *av* outcomes, in which the *ap* outcomes dominate. Between the midpoint and the *av* pole lies an array of points representing types of conflict in which the *av* outcomes dominate.

Another and theoretically more parsimonious way of characterizing this dimension is in terms of risk (see Fig. 1). Risk may be defined as the relative magnitude of the total *ap* and *av* tendencies present in the situation of choice. The degree of risk will be a direct function of the above ratio. The *ap* and *av* poles (±1.00) represent the minimal risk ends of the continuum. The midpoint, 0.00, represents the point of maximal risk. In other words, types of conflict may be represented as varying degrees of risk.

Six different types of conflict, or situations of risk, were examined in this paper. They may be designated as follows: (1) the double *ap-ap* type; (2) the double *av-av* type; (3) a type consisting of an *ap-ap* and an *ap-av* alternative; (4) a type consisting of an *av-av* and an *av-ap* alternative; (5) the double *ap-av* type, where the gain was always larger than the loss; (6) the double *av-ap* type, where the loss was always larger.

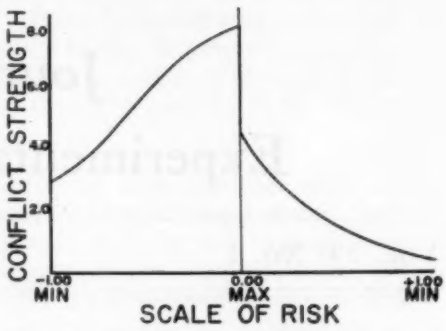


FIG. 1. Theoretical relationship between risk (type of conflict) and conflict strength.

These six types may be reduced to three major types of conflict (see Table 1). Each would be represented by an *ap*-dominant (possible gains are always greater) and an *av*-dominant (possible losses are always greater) subtype. Situations 1 and 2 are both characterized by uniformity of reinforcement associated with the outcomes. All of the outcomes are approximately equivalent for both subtypes, except for the possibility of winning in Situation 1 and losing in Situation 2. Situations 1 and 2 have been labeled the *univalent* conflict (Type I) and represent extensions of the familiar *ap-ap* and *av-av* types. Conflict Situations 3 and 4 are distinguishable by the dissimilarity of the alternatives. In both subtypes, a "gamble" is opposed to a "sure thing." The alternatives are approximately the same in each subtype but reversed in *ap* and *av* designation. Because of the dissimilarity between the competing alterna-

[2] As long as *S* is forced to relinquish the alternative not chosen, a pure *ap-ap* or *av-av* type conflict would not exist. There is always the possibility of making a better (*ap*) or worse (*av*) choice. If *S* knows he can simultaneously obtain both alternatives, there would be no risk present and, hence, no conflict.

TABLE 1

TYPES OF CONFLICT

(*ap* SUBTYPES ARE PRESENTED TO THE LEFT AND *av* SUBTYPES TO THE RIGHT)

Type I. Univalence, minimal risk

| | A | B | | A | B |
|---|---|---|---|---|---|
| a | win 35¢ | win 30¢ | a | lose 34¢ | lose 30¢ |
| b | win 26¢ | win *x*¢ | b | lose 25¢ | lose *x*¢ |

Type II. Disvalence

| | A | B | | A | B |
|---|---|---|---|---|---|
| a | win 35¢ | win 5¢ | a | win 26¢ | lose 5¢ |
| b | lose *x*¢ | win 5¢ | b | lose *x*¢ | lose 5¢ |

Type III. Ambivalence, maximal risk

| | A | B | | A | B |
|---|---|---|---|---|---|
| a | win 34¢ | win 26¢ | a | win 25¢ | win 17¢ |
| b | lose 25¢ | lose *x*¢ | b | lose 34¢ | lose *x*¢ |

Note.—A and B are alternative choices; a and b represent possible outcomes. The probability of occurrence of any a or b was equal to one-half. *x* represents the unique value producing indifference of choice between A and B.

tives, these two subtypes have been labeled the *disvalent* conflict (Type II). Conflict Situations 5 and 6 are subtypes of the familiar *ambivalent*, or double *ap-av*, conflict. The expected gain is always larger in Situation 5 and smaller in Situation 6. The outcomes are the same in both subtypes, but the *ap* and *av* signs are reversed. The three major conflict types are assumed to be represented along the dimension of risk such that their subtypes are symmetrically distributed about the midpoint (see Fig. 2). The greatest amount of risk would attend the ambivalent conflict and the least amount the univalent conflict.

It is generally concluded that *av* situations yield stronger and more persistent responses than equivalent *ap* situations (Atthowe, Davidson, Siegel, & Suppes, 1954; Barker, 1942; Hovland & Sears, 1938; Lewin, 1935; McClelland, 1951; Miller, 1944; Nogee, 1953). Therefore, we should expect the function representing the strength of a conflict to be higher on the *av* side of the continuum. Figure 1 represents the theoretically expected relationship between risk (the dimension underlying types of conflict) and conflict strength. The shape of the curve must take into account the difference between *ap* and *av* instigated behavior and represent a nonmonotonic function.

The present study divides into three problems, each identified with a measure of conflict resolution. The first problem is concerned with the effect of differing degrees of risk upon the latency of a response. Response latency was defined as the time it takes *S* to make a decision. Decision time has been shown to be one of the most "sensitive" measures of choice or judgment (Cartwright, 1941). In general, as two objects or events approach "equivalence," the longer becomes the decision time (Deutsch, 1954; Johnson, 1945). Decision time has also been related to the degree of confidence in one's decision (Johnson, 1945) and to emotional behavior (e.g., complex detection [Woodworth, 1938]).

It is hypothesized that decision time as an index of conflict strength will be related to risk, or type of conflict, as represented by the theoretical curve in Fig. 1. The following predictions would, thereby, be generated:

*Hypothesis I.* The three major types of conflict will differ significantly in decision time.

*Corollary 1a.* The ambivalent conflict will show the longest decision times.

*Corollary 1b.* The univalent conflict will show the shortest decision times.

*Hypothesis II.* The *av*-dominant subtypes will show longer decision times than the *ap*-dominant subtypes.

In addition, we are concerned with the effect of risk upon the frequency and degree of nonrational responses. Choices were designated as either rational or nonrational. A choice equivalent to that made by a mathematical model, or a purely "rational man," was scored as rational; any choice differing from this choice was termed nonrational. The larger the difference between the observed choice and rational choice, the greater would be the degree of nonrationality. The meaning of a rational or nonrational response is unclear. Any definition must at present be arbitrary; nevertheless, a criterion of "maximization" is probably essential. Nonrational, as well as rational, behavior is certainly found in situations involving conflict, but there is little evidence relating degree of nonrationality to the type of conflict. Is nonrationality a measure of conflict strength?

The following general questions were asked in regard to the rationality or nonrationality of a decision conflict: (1) What is the relationship between risk (type of conflict) and the degree of rationality of a decision? (2) What is the relationship between rationality and decision time?

The third problem dealt with *S*s' strategy, or pattern of responses, and the relationship of strategy to risk. Two general types of nonrational strategies were identified together with the "completely rational" strategy. The two nonrational strategies were: (*a*) *conservatism*, an overconcern with the possibility of losing money, and (*b*) *extravagance*, an overconcern with the possibility of gaining money. An *S* who more frequently chose the alternative representing the least possibility of losing rather than the alternative representing the greatest possibility of winning was termed a conservative strategist. If the latter alternative was chosen more frequently, *S* was labeled extravagant. For example, in deciding between Alternatives A and B below, a conservative would choose B, and an extravagant *S* would choose A.

| | A | B | |
|---|---|---|---|
| (*a*) | win 35¢ | win 26¢ | $P = \frac{1}{2}$ |
| (*b*) | lose 26¢ | lose 17¢ | $P = \frac{1}{2}$ |

A completely rational *S* would indicate that both alternatives were equivalent in expected value.

Game theorists (e.g., Dunlap, 1955; Luce & Raiffa, 1957; Thrall, Coombs, & Davis, 1954) have introduced the concept of strategy into experimental psychology. Their emphasis has been upon the best type of strategy. The emphasis in this paper is upon the types of strategy typically used and the conditions associated with a given strategy. Specifically, the questions asked were the following: (1) What are the typical types of strategies used in resolving conflicts? (2) What is the relationship between risk and the strategy used?

## METHOD

*Task.*—A simple yet fairly representative task involving a series of choices was developed. The task, a gambling game, was both interesting and motivating. The instructions were designed to induce moderate to strong achievement motivation and a reasoning attitude. The alternative choices were represented as "wagers." They were said to be mathematical reasoning problems taken from an advanced intelligence test. The *S*s were told that the objective of the investigation was to study the influence of a gaming context upon general reasoning ability.

The *S*s were instructed to choose one of two alternative wagers. Each alternative consisted of two possible outcomes (see Table 1). The probability of either outcome was one-half. The outcomes symbolically represented differing monetary values, although no money was actually involved. In making his choice, *S* had to weigh the possible numerical consequences (i.e., outcomes) of each alternative before making his decision. Because no money was gained or lost and *S*s were never told whether their decisions were correct, the effects of reinforcement were assumed to be nullified.

The event determining the occurrence of either outcome was stated to be the rolling of a die. If the die rolled an even number, Outcome a was said to occur (see Table 1); otherwise Outcome b occurred. The actual rolling of the die was never performed. The die was simply used as a prop.

*Subjects.*—The *S*s were 32 male students attending a class in introductory psychology at Stanford University. They were each tested individually for approximately 1.5 hr.

*Procedure.*—Each of the six conflict types was randomly presented in the series of choices. In each type, the decision analyzed was that made at the point of conflict. The point of conflict was taken as the point of indifference between the two alternatives (i.e., the PSE). The indifference point was established by the method of constant stimuli. The outcomes of one alternative (e.g., Alternative A in the *ap* subtype, ambivalent conflict, Table 1) were held constant throughout the session; whereas one of the outcomes (Outcome b in Alternative B) of the other alternative was varied. This outcome, designated by $x$, was varied until the two alternatives were reacted to as indifferent. For example, if $x$ in the above subtype assumed at random the values 10, 14, 16, 18, 20, and 24, and *S*'s choices were B, B, B, A, A, and A, then the point of indifference lies midway in the interval bounded by 16 and 18 (i.e., 17).

The time required to make a choice and the particular choice were both determined at the point of indifference. Decision time was measured in units of .1 sec. with an electric timer. A rational, or mathematically correct, choice was operationally defined as that choice in which the mathematical expected values of the alternatives were equal in magnitude. In the above example, 17 is also the rational choice. Any choice corresponding to an indifference point which deviated from this value was termed nonrational. In the above example, if *S* had preferred B for the values of 10, 14, 16, and 18 and A for 20 and 24, then his indifference point would be 19. In this case, the degree of nonrationality represented by 19 would be 2 (i.e., 2 cents or units away from a rational choice).

The *S* was separated from *E* by a low screen. The wagers, printed on $3 \times 5$ cards, were presented to *S* through a small opening at the bottom of the screen. As the card emerged on *S*'s side of the screen, the timer was started. The *S* was instructed to make only one choice. An *O*, seated a few feet away, recorded choices and decision times. The *E* also kept a tally of *S*'s choices in order to determine *S*'s point of indifference. If 12 choices were made within a given subtype without the determination of an indifference point, *S* was automatically discarded. Seven *S*s were discarded for this reason.

## Results

*Decision time.*—The distribution of decision times was markedly skewed. Further, Bartlett's test (Edwards, 1950) for homogeneity of variance was rejected at the .001 level. Neither logarithmic nor reciprocal transformations increased the homogeneity of variance. Therefore, ordinal measurement was assumed, and distribution-free methods were used to analyze the data.

Hypothesis I was confirmed. The six subtypes as well as the three major types of conflict yielded significantly different decision times ($P \leq .001$ in both instances). The test of significance used was Friedman's two-way analysis of variance of ranked data (Siegel, 1956).

Corollary 1a was confirmed. The difference between the ambivalent and disvalent types (the smallest difference) was determined for each *S* and then ranked in magnitude. The ambivalent type of conflict yielded significantly longer decision times ($P \leq .025$). The Wilcoxon Signed-Ranks test (Siegel, 1956) was used to test the difference.

TABLE 2

CONFLICT RESOLUTION AS A FUNCTION OF TYPE OF CONFLICT

| Type of Conflict | Nonrationality | | Decision Times (Sec.) | |
|---|---|---|---|---|
| | Freq. $N = 26$ | Degree $N = 19$ | Mean $N = 19$ | Rank $N = 26$ |
| Ambivalence | | | | |
| *ap* | 17 | 36 | 21.8 | 3.3 |
| *av* | 15 | 20 | 27.9 | 5.1 |
| Disvalence | | | | |
| *ap* | 23 | 135 | 20.7 | 3.0 |
| *av* | 25 | 183 | 25.0 | 4.2 |
| Univalence | | | | |
| *ap* | 12 | 30 | 18.1 | 2.6 |
| *av* | 10 | 16 | 19.8 | 2.9 |

Note.—Rank represents the mean of the rankings for a given type of conflict. A rank of 6.0 equals the longest decision time.

Corollary 1b was confirmed. The difference between the univalent and disvalent types (the smallest difference) was determined. The univalent conflict yielded significantly shorter decision times ($P \leq .05$, Wilcoxon test).

Hypothesis II was confirmed. The *av* subtypes showed significantly longer decision times than the *ap* subtypes ($P \leq .01$, Wilcoxon test).

The results are presented in tabular form in Table 2. Six *S*s showed a rational strategy (i.e., showed no nonrational choices). Consequently, they did not enter into the computations; nevertheless, their responses as a group served to clarify the relationship between risk and decision time. The six rational *S*s also showed significant differences in decision time as a function of the type of conflict ($P \leq .05$, Friedman test, $df = 5$). The ambivalent conflict yielded the longest decision times, but the disvalent conflict showed the shortest decision times. All statistical tests reported in this section were one-tailed tests, except the latter.

*Nonrationality.*—Both the frequency and the degree of nonrational choices were found to be significantly different among the subtypes and between the three major types of conflict ($P \leq .001$). Cochran's $Q$ test (Siegel, 1956) was used to evaluate frequency and Friedman's test to evaluate degree. The disvalent conflict clearly yielded the largest frequency ($Q$ test) and the greatest degree (Wilcoxon test) of nonrationality when compared to either of the other major types ($P \leq .001$). The univalent conflict generally exhibited the fewest number ($Q$ test) and the least degree (Mood's Chi-square test for correlated measures [McNemar, 1955]); the latter trends approached but did not reach significance ($P \leq .15$ in both instances). Because the differences were often similar in magnitude, Mood's test was considered more appropriate than that of Wilcoxon.

No significant difference was shown between the *av* and the equivalent *ap* subtypes, either in terms of the frequency or the degree of nonrationality. The rank-order correlation (rho) between the degree of nonrationality exhibited by *S* and his median decision time was .04. This independence between the rationality and the latency

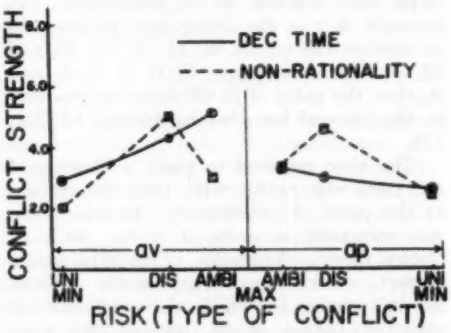


FIG. 2. Conflict strength as a function of risk. One scalar unit of strength represents one unit of decision time (ranks) and five units of nonrationality (frequency).

of a decision is further shown in Fig. 2. Whereas the relationship between risk and decision time is monotonically increasing, the relationship between risk and nonrationality was nonmonotonic. All statistical tests involving nonrationality were two-tailed tests.

*Strategy.*—Of 18 *S*s who were classified as having dominant nonrational strategies, 83% were conservative. This difference was significantly different from chance expectations at the .01 level, as determined by chi square. If all nonrational choices are considered, 71% were in a conservative direction. As would be expected, all 13 *S*s who consistently refused to take a chance (i.e., gamble) in the disvalent conflict used a conservative strategy ($P \leq .001$, chi square, when compared with chance expectations). The one *S* who consistently took a chance used an extravagant strategy. Four *S*s varied their strategies from situation to situation.

An inspection of Table 3 indicates that more rational decisions were made in the univalent situation; no distinction was made between the nonrational strategies. The strategy exhibited in the disvalent conflict was predominantly conservative. In the ambivalent conflict, very few decisions were extravagant; both conservative and rational strategies were more frequently manifested. If the total *N* is inflated by assuming that the observations in each choice situation were made independently, a chi square can be computed. The overall and pairwise tests of significance (two-tailed) were beyond the .001 and .01 levels when the three types of conflicts were compared in terms strategy.

TABLE 3

STRATEGY AS A FUNCTION OF TYPE OF CONFLICT

| Strategy | Type of Conflict | | | | | | Total |
|---|---|---|---|---|---|---|---|
| | Ambivalent | | Disvalent | | Univalent | | |
| | $a_v$ | $a_p$ | $a_v$ | $a_p$ | $a_v$ | $a_p$ | |
| Conservative | 10 | 14 | 20 | 17 | 4 | 7 | 72 |
| Extravagant | 5 | 3 | 5 | 6 | 6 | 5 | 30 |
| Rational | 17 | 15 | 7 | 9 | 22 | 20 | 90 |
| Total | 32 | 32 | 32 | 32 | 32 | 32 | 192 |

Note.—The tabled values refer to the number of *S*s exhibiting a given strategy.

## DISCUSSION

The results indicate the necessity of considering the type of conflict, or risk, in predicting decision making behavior. By equating the outcomes along four dimensions, differences in decision making were assumed to be a function of the situational arrangement of the outcomes. The subjective probability of the occurrence of any outcome, the relative magnitude of their difference within an alternative, the absolute expected value (mathematical) of the sum of the outcomes per alternative, and the range of outcome values were held constant or relatively so. Table 1 shows that all of these controls could not be carried out simultaneously. The expected values of the univalent alternatives were approximately equal to 30; those of the other alternatives were approximately 5. The net effect of this discrepancy would be to overweight the univalent conflict and reduce the likelihood of reaching significance; consequently, the substantiation of the hypotheses may be viewed with greater certainty.

The relative balance of $a_p$ and $a_v$ tendencies (i.e., risk) was shown to be directly related to the strength of a conflict as measured by decision time. The greater the risk, the longer the decision time. The relationship between risk and nonrationality was not as clear. The results indicate that a moderate amount of risk was associated with

maximum nonrationality, although the conflict representing moderate risk (disvalence) was also characterized by the option of gambling. It seems more likely that the instrumental uncertainty (whether to gamble or not) involved in making a decision is more important in determining its correctness. This latter interpretation is suggested by the comparison of the completely "rational" with the nonrational *S*s. Both groups exhibited longer decision times in the ambivalent conflict, but the rational *S*s' shortest latencies were in the disvalent conflict, while those of the nonrational *S*s were in the univalent conflict. Risk alone cannot account for this difference. It is suggested that a concomitant of rationality is a reduction in instrumental uncertainty, in this case the concern with taking chances.

The *S*s tended to minimize their maximum losses rather than maximizing their maximum gain. These findings lend empirical support to the "minimax" principles found in game theory (e.g., Luce & Raiffa, 1957); although, as Mosteller and Nogee (1951) have shown, college students are probably a rather conservative sample. In addition, the type of conflict or risk provides a basis for differences in strategy. The *S*s tended to be significantly more rational when risk and uncertainty were minimal. As risk and instrumental uncertainty increased, conservatism also increased. The option of gambling not only was associated with more nonrational decisions but also with much greater conservatism.

The equation of types of conflict and risk provides a basis for a clearer and more formal exposition of conflict theory. The focus upon anticipated outcomes also suggests an interrelationship between decision making (judgmental processes) and conflict (motivational processes).

The gaming paradigm has a far wider experimental application than in the testing of statistical models, per se. Its appeal to many diverse populations (e.g., Hurst & Siegel, 1956; Mosteller & Nogee, 1951) plus its simplicity provides a ready experimental context in which to test and to compare judgmental and motivational conditions. In addition, the method of indifference-point analysis yields a basis for equating and comparing individual reactions and situations of conflict.

## Summary

The relationship between decision-making behavior and differing types of conflict was investigated. The experimental task involved choosing the "better" of two alternative wagers. Conflict was defined in terms of the arrangement of the expected outcomes. Risk, the relative magnitudes of *ap* and *av* outcomes, was postulated to be one of the major dimensions underlying types of conflict. As risk increased, the conflict strength, as defined by decision time, also increased. Rationality was associated with minimal risk, whereas nonrationality was mainly associated with uncertainty (chance taking). Decision time and the rationality of a decision were found to be independent measures of conflict resolution. The *S*s tended to be more conservative, especially as risk and uncertainty increased.

## References

Atthowe, J. M., Jr., Davidson, D., Siegel, S., & Suppes, P. An experimental study of some factors influencing decision-making in conflict situations involving risk. Paper read at West. Psychol. Ass., Long Beach, April, 1954.

Barker, R. G. An experimental study of the resolution of conflict by children. In Q. McNemar & M. K. Merrill (Eds.), *Studies in personality*. New York: McGraw-Hill, 1942. Pp. 13–34.

Cartwright, D. Decision-time in relation to the differentiation of the phenomenal field. *Psychol. Rev.*, 1941, **48**, 425–442.

Deutsch, M. Field theory in social psychology. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 181–222.

Dunlap, J. W. *Mathematical models for human behavior*. Stamford, Conn.: Dunlap Assoc., 1955.

Edwards, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

Hovland, C. I., & Sears, R. R. Experiments on motor conflict: I. Types of

conflict and their mode of resolution. *J. exp. Psychol.*, 1938, **23**, 477–493.

HURST, P. M., & SIEGEL, S. Prediction of decisions from a higher ordered metric scale of utility. *J. exp. Psychol.*, 1956, **52**, 138–144.

JOHNSON, D. M. A systematic treatment of judgment. *Psychol. Bull.*, 1945, **42**, 193–224.

LEWIN, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.

LUCE, R. D., & RAIFFA, H. *Games and decisions: Introduction and critical survey.* New York: Wiley, 1957.

MCCLELLAND, D. C. *Personality.* New York: Sloane, 1951.

MCNEMAR, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.

MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. I. New York: Ronald, 1944. Pp. 431–465.

MOSTELLER, F., & NOGEE, P. An experimental measurement of utility. *J. polit. Econ.*, 1951, **59**, 371–404.

NOGEE, P. An experiment in decision-making under uncertainty. Paper read at the Seminar in the Application of Mathematics to the Social Sciences. Univer. of Michigan, November, 1953.

SIEGEL, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

THRALL, R. M., COOMBS, C. H., & DAVIS, R. L. (Eds.), *Decision processes.* New York: Wiley, 1954.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received March 3, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 1, 1960

# AGE DIFFERENCES IN CARD-SORTING PERFORMANCE IN RELATION TO TASK DIFFICULTY, TASK SET, AND PRACTICE

JACK BOTWINICK, JOSEPH S. ROBBIN, AND JOSEPH F. BRINLEY[1]

*Laboratory of Psychology, National Institute of Mental Health*

Inferences about aging processes may be made from functional relationships between task difficulty and age differences in task performance. For example, from such relationships, Welford (1958) has suggested that information capacity of the sensorimotor system is lowered with age, Botwinick, Brinley, and Robbin (1958a) have suggested that with age there may be an increase in response reviewing tendencies, and Birren and Botwinick (1955) have emphasized perceptual changes with age.

In the present study, performances of two age groups were compared when task difficulty was defined in two different ways but when the task itself, card sorting, was kept the same. The problem of task set was also studied. These three aspects of the task were investigated with respect to age differences in the effects of practice.

## METHOD

*Subjects.*—There were two age groups of male volunteer *Ss*. The older group comprised 27 *Ss* aged 65 to 81 yr. with a median of 71.0 yr., and the younger group comprised 30 *Ss* aged 19 to 35 yr. with a median of 24.5 yr.

Both age groups were divided into high- and low-education subgroups—those with 12 or more years of formal education and those with less than 12 years. In the older group, 14 *Ss* were in the high-education subgroup and 13 were in the low- subgroup. The range of education was 4 to 11 and 12 to 16 yr. in low- and high- subgroups respectively. All elderly *Ss*, except one, were community residing.

The high-education subgroup of young *Ss* comprised 10 religious research volunteers and 8 Public Health Service (PHS) employees. The low- subgroup of younger *Ss* comprised 12 workers of the PHS. The range of education was 6 to 11, and 12 to 18 yr. in low- and high- subgroups, respectively.

*Procedure.*—Each *S* performed in a series of three experiments that involved card-sorting tasks. The tasks were completed during the course of three sessions which lasted a total of approximately 3.5 hr. Two younger *Ss*, however, completed the tasks in fewer sessions. The *Ss* held 71 playing cards face up and sorted as quickly as possible into slots according to instructions. The 71 cards were of two complete decks from which the 24 face cards and the 9 stimulus cards shown in Table 1 were removed. Each card was sorted into a 3.6 × 5.0-in. slot of a slightly tilted sorting board 36 in. in the left to right dimension, divided equally into 10 slots. When a card was sorted, it slid out of sight through a narrow opening at the back end of the slot. Placed perpendicular to the slot and facing *S*, was the stimulus card that indicated the sort.

*Experiment I.*—Each *S* sorted the 71 cards by matching the number on the card with the number on the stimulus card. When no match was possible because none of the stimulus numbers corresponded to the sorting card, it was placed in the tenth slot, that is the last slot, which was designated "bin."

There were five levels of card-sorting difficulty defined by the number of stimulus cards or slots that were sorted. The difficulty levels were 2, 4, 6, 8, and 10 slots. Since a different number of cards were sorted into the bin for each level of difficulty, the abscissa in Fig. 1 is labeled a, b, c, d, and e, representing lesser to greater task difficulty. As indicated in Table 1, all 36 in. of the sorting board were used for each level of difficulty. Each of the five different levels was sorted four consecutive times, i.e., 4 trials. There was a shuffling of the cards between each trial. Difficulty levels were presented in random order except that Level c was always last.

[1] We thank James E. Birren, Chief, Section on Aging, NIMH, for suggestions relating to this study.

TABLE 1

ARRANGEMENT OF STIMULUS CARDS ON SORTING BOARD FOR VARIOUS TASK DIFFICULTY LEVELS

| Task | Sorting Board Slot Numbers | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 (bin) |
| Warm up | Q | | | J | | | K | | | X |
| Level a | 3C | | | | | | | | | X |
| Level b | 7C | | | 4D | | | 9C | | | X |
| Level c | 5H | | 9C | | 3C | | 8S | | 6S | X |
| Level d | 7C | | 4D | 10S | 6S | 3C | 5H | 2D | | X |
| Level e | 4D | 10S | 6S | 9C | 2D | 5H | 7C | 3C | 8S | X |

Note.—Stimulus cards are shown in the position they occupied on the sorting board for the various levels of task difficulty. Thus, for difficulty Level a, Stimulus Card 3C occupied the first slot while for Level c, it occupied the fifth slot. All sort cards that could not be placed properly in Slots 1 through 9, were placed in tenth slot (bin). Q, J, K = face cards, C = clubs, D = diamonds, H = hearts, S = spades, and X = extra sort cards.

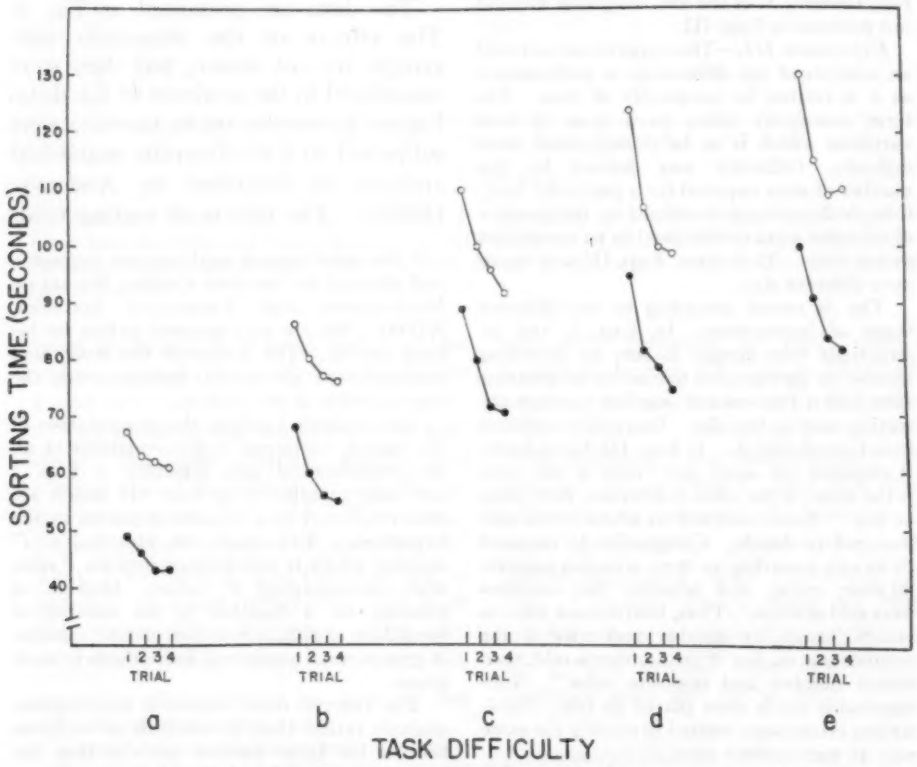


FIG. 1. Mean time to sort 71 cards by matching sort and stimulus card numbers. Abscissa refers to number of slots or match numbers, with difficulty Level a being the fewest and Level e being the largest number of slots. Each difficulty level was sorted for four trials by *Ss* 65 to 81 yr. (open circles), and by *Ss* 19 to 35 yr. (filled circles).

This was for a purpose that will be described later.

All *Ss* were given warm-up practice before the experiment began. The warm-up involved three trials of matching face cards shown in Table 1. Each trial comprised 21 face cards, plus 8 sevens and 8 aces for sorting in bin.

*Experiment II.*—This experiment included an analysis of age differences in the effects of changing the task set or card-sorting habit. Immediately following the four trials of sorting with the 6-slot difficulty level (Level c), an additional four trials were given followed by a different arrangement of the same stimulus cards. The new arrangement was 9C, 6S, 8S, 5H, and 3C, and bin. Four consecutive trials were given with the new stimulus arrangement. The *Ss* for this experiment were all those of the young group, 6 elderly of the high-education subgroup, and 10 elderly of the low- subgroup plus 3 additional elderly *Ss* of the low- subgroup who did not perform in Exp. III.

*Experiment III.*—This experiment included an analysis of age differences in performance as it is related to complexity of task. The term *complexity* refers to a type of task variation which is to be distinguished from *difficulty*. Difficulty was defined by the number of slots required for a particular task. Complexity of task is defined by the number of stimulus aspects that need to be considered at one time. Most often, Exp. III was begun on a different day.

The *Ss* sorted according to two different types of instructions. In Exp. I, the instructions were simply to sort by matching number of sorting card to number of stimulus card, and if this was not possible, to place the sorting card in the bin. This will be referred to as Complexity A. In Exp. III, Complexity B required the same sort "only if the color is the same; if the color is different, then place in bin." Color referred to whether the suit was red or black. Complexity C required *Ss* to sort according to three stimulus aspects: number, color, and whether the numbers were odd or even. Thus, instructions were to match "cards by number and color if the number is even, but if the number is odd, then match number and opposite color." Non-matchable cards were placed in bin. Card-sorting errors were treated in exactly the same way as were correct sorts.

Sorting was not begun until both *S* and *E* were satisfied that the instructions were understood. Instructions were repeated before each trial and as often as necessary. Within Complexities B and C, there were two levels of task difficulty: 4 and 10 slots (Levels b and e). Orders of Complexity B and C, and Difficulty Levels b and e were given for four trials each in random order. The same warm-up for sorting was given before Tasks B and C as was given before Task A. Tasks B and C were given most often in two sessions and nearly always in two separate days. In the 4-slot difficulty level for both Complexities B and C, the stimulus card 9H was substituted for the 9C on alternate trials in order to discourage correct sorting by a different method that was found to be present with early *Ss*. Three elderly *Ss* of the low-education subgroup did not perform in Exp. II Their scores, however, were not different from the mean of the subgroup.

## Results

### *Experiment I*

The data are presented as Fig. 1. The effects of the education subgroups are not shown, but they were considered in the analyses of the data. Except for sorting trials, the data were subjected to a multivariate statistical analysis as described by Anderson (1958).[2] The effects of sorting trials

[2] The multivariate analysis was suggested and directed by Seymour Geisser, Section on Mathematics and Theoretical Statistics, NIMH. We are very grateful to him for his kind service. The design of the study and interpretation of results, however, are the responsibility of the authors.

The analysis involves the computation of the means, variances, and correlations of all the conditions of task difficulty in Exp. I and task complexity in Exp. III which are then combined in a manner requisite to the hypothesis. This leads to Hotelling's $T^2$ statistic which is transformed into an $F$ ratio with corresponding $P$ values. Degrees of freedom are a function of the number of conditions of difficulty or complexity, number of groups, and number of individuals in each group.

The current data require a multivariate analysis rather than an analysis of variance because the latter method assumes that the variation of individuals be the same among the different conditions of difficulty or complexity, and that the conditions be uncorrelated. These assumptions are rarely met when replicate observations are made on the

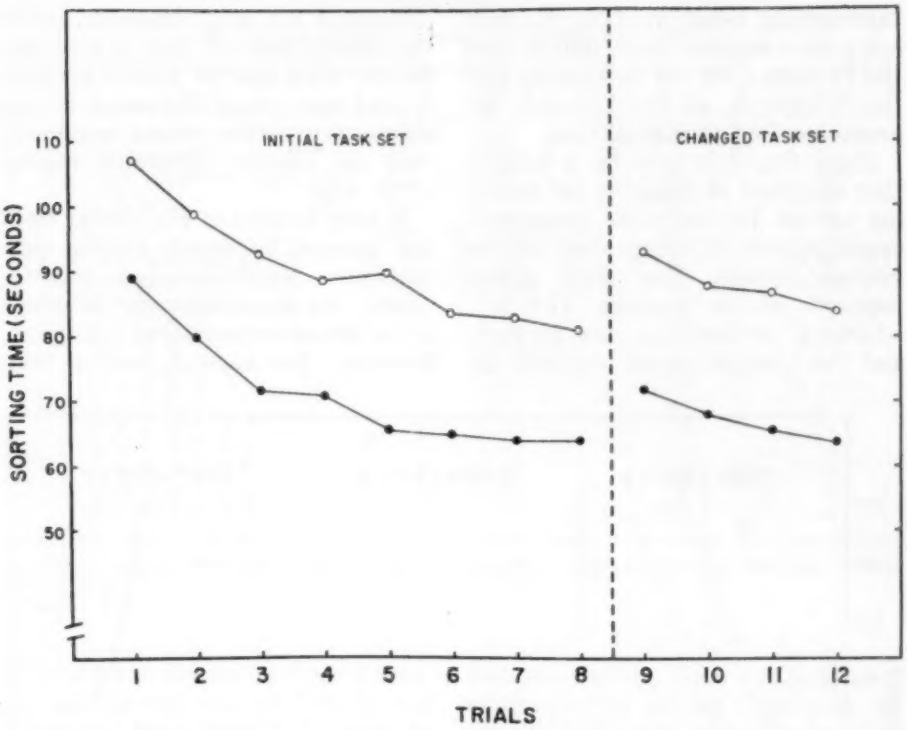


FIG. 2. Mean time to sort 71 cards by matching sort and stimulus card numbers. Each sort was done for eight consecutive trials followed by an additional four trials with a rearranged stimulus pattern. The *S*s were 65 to 81 yr. (open circles), and 19 to 35 yr. (filled circles).

were analyzed by *t* tests utilizing the values associated only with Trials 1 and 4.[3]

The older group was found to sort slower than the younger group, and sorting speed decreased in relation to increased task difficulty. There was improvement in speed of sorting from Trial 1 to Trial 4, and this tended, in general, to occur at a faster rate with the more difficult tasks. The two education subgroups were not significantly different in their sorting speeds except in relation to the interaction of age and task difficulty. This relation (Education × Age × Task difficulty) was unsystematic and difficult to interpret, but the low-education subgroup of younger *S*s was consistently poorer than the better educated *S*s of the same age. The *F* ratios which refer to these results were statistically significant ($P < .01$), but all other tests of the analyses were not ($P > .05$).

*Experiment II*

The data of this experiment are presented as Fig. 2. It may be seen that with the change of task set or

same *S* in all treatments, i.e., levels of task difficulty or complexity.

[3] *t* tests were performed when the requirements of multivariate analysis were not met in the data or when only two conditions were involved. Variance analyses were not performed because of disproportionate subgroup frequencies.

card-sorting habit in Trial 9, both age groups required more time to sort the 71 cards. Of the 30 younger and the 19 older *S*s, all but 3 in each age group required additional time.

From Fig. 2 it may be computed that the effect of changing the stimulus pattern for the older group was approximately 1.5 times that for the younger group. The older group required, on the average, 11.9 sec. additional to sort the new pattern, and the younger group required an additional 8.0 sec. However, when the sorting time of Trial 8 was subtracted from that of Trial 9 for each *S*, and age groups compared in this difference in sorting time, it was found that no reliable difference existed ($P > .05$).

It may be seen in Fig. 2 that both age groups improved sorting performances with increased practice trials. No appreciable age difference in the rate of improvement was found, however. For each *S*, sorting time

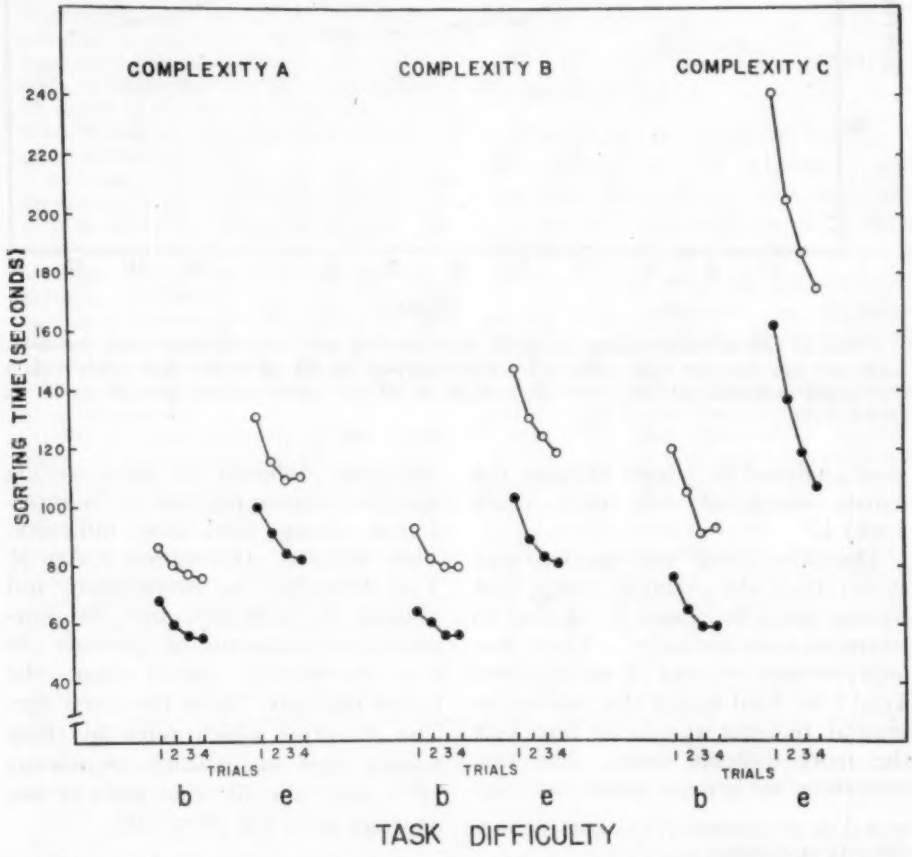


FIG. 3. Time to sort 71 cards in relation to task difficulty (number of stimulus cards matched), and task complexity (number of stimulus aspects kept in mind and manipulated simultaneously). There were four trials for each level of task complexity and difficulty. Task Level A and b involved the fewest number of stimulus cards and stimulus aspects, and Task Level C and e the largest number. The *S*s were 65 to 81 yr. (open circles), and 19 to 35 yr. (filled circles).

TABLE 2

MULTIVARIATE ANALYSIS OF DATA FROM EXP. III

| Source | *df*[a] | *T*[a] | *F* |
|---|---|---|---|
| Age (A) | 3, 49 | 38.76 | 12.43** |
| Education (E) | 3, 49 | 2.43 | .78 |
| Task complexity (C) | 2, 52 | 10.64 | 5.22** |
| A × C | 2, 52 | 10.46 | 5.14** |
| A × E | 3, 49 | 1.59 | .51 |
| E × C | 2, 52 | 1.71 | .84 |
| Task difficulty (D) × C | 2, 52 | 131.61 | 64.56** |
| Sorting trials (T) × C | 2, 52 | 29.08 | 14.27** |
| A × E × C | 2, 52 | .42 | 2.06 |
| A × C × D | 2, 52 | 9.73 | 4.77* |
| A × C × T | 2, 52 | 2.31 | 1.13 |
| D × C × E | 2, 52 | .08 | .04 |
| D × C × T | 2, 52 | 11.19 | 5.49** |
| T × C × E | 2, 52 | .97 | .47 |
| A × D × C × T | 2, 52 | .08 | .04 |
| A × C × T × E | 2, 52 | 1.10 | .54 |
| A × D × C × E | 2, 52 | 3.42 | 1.68 |
| D × C × T × E | 2, 52 | .25 | .12 |
| A × D × C × E × T | 2, 52 | 6.35 | 3.12 |

[a] The *df*'s are not determined in the same way as in analysis of variance. They are based upon different statistical operations.
* $P < .05$.
** $P < .01$.

of Trial 8 was subtracted from Trial 1, and also sorting time of Trial 12 was subtracted from Trial 9, in order to obtain measures of improvement in both the original sorting task and in the task of changed stimulus pattern. No significant differences between age groups were found with respect to these measures ($P > .05$).

*Experiment III*

Card-sorting time was analyzed in relation to age, education, task difficulty, task complexity, sorting trials, and the interactions among these conditions. The data, except for the effects of education, may be seen in Fig. 3. Relations were evaluated statistically by multivariate and *t* test analyses, as before, and the results of these tests are shown in Tables 2 and 3.

The multivariate analysis indicated that the effect of increased age and increased task complexity was to increase sorting time. In addition, relative to the sorting times with the simple task, the older group was slower with increased complexity of task than was the younger group. All three differences (age, task com-

TABLE

COMPARISONS OF MEANS BY *t* TEST OF DATA FROM EXP. III

| Source[a] | *t* |
|---|---|
| Trials (T) | 16.48** |
| Task difficulty (D) | 4.44** |
| Age (A) × T | 1.51 |
| A × D | 3.79** |
| T × D | 2.06* |
| Education (E) × T | 1.46 |
| E × D | .46 |
| A × T × E | 1.40 |
| A × D × E | .77 |
| A × D × T | .50 |
| D × T × E | 1.09 |
| A × D × T × E | 1.96 |

[a] *df*'s for all sources = 53 (the sum of the four subgroup *df*'s). Analyses of variance not performed because of disproportionate subgroups *N*s.
* $P < .05$.
** $P < .01$.

plexity, and the interaction between age and task complexity) were significant ($P < .01$). It may be seen in Fig. 3 that sorting time was longest with the task that was a combination of most difficult and most complex ($P < .01$). This effect of combined task difficulty and complexity was disproportionately greater for the older group ($P < .03$).

The effects of sorting trials were investigated by use of Trials 1 and 4. Rate of improvement with practice was positively related to increased task complexity and to the interaction of increased task complexity and difficulty ($P < .01$). The data involving Sorting Trials 1 and 4 as related to age, education, and task difficulty were subjected to a series of $t$ tests. The effect of practice was to decrease sorting time, the effect of increased task difficulty was to increase sorting time, and the improvement with practice occurred at a faster rate with the more difficult problems. Unlike the statistical results of Exp. I, the Age × Difficulty interaction was significant, indicating that the effect of increased task difficulty was to increase sorting time disproportionately for the older group. These differences, except for the Trials × Difficulty interaction, were all significant at $P < .01$; the Trials × Difficulty interaction was significant at $P < .05$. As with task complexity, the effect of trials was similar for both age and education groups, and all the second-order interactions involving education were nonsignificant ($P > .05$).

## Discussion

There are three main aspects to the results of these studies. The first is that increased task difficulty and complexity tends to slow the elderly more than the young. The second aspect is that while there is improvement with continued sorting, and while rate of improvement tends to increase with increased task difficulty and complexity, improvement rate is about the same for the elderly as for the young *S*s. The third aspect is that high- and low-education groups were not significantly different in sorting speed.

It is important to evaluate the task variations within the three experiments with respect to age. The motor or movement component of the various card-sorting performances was more or less constant from one experimental condition to another. Thus the effects of task difficulty (Exp. I), task set (Exp. II), and task complexity (Exp. III) were primarily on nonmotor aspects of card sorting. The increase in sorting time with increased task difficulty was due, therefore, more to the perceptual operations of searching for or matching of the specific card number, or to the selection and programming of responses, than to motor speed. The effect of age on this function was not appreciable. The effect of age was more apparent when the task was varied in complexity by the number of stimulus aspects that needed to be kept in mind and manipulated simultaneously. The relative increase in sorting time by the elderly with increased complexity was taken as an index of age deficit in this type of cognitive function rather than in motor speed. The number of stimulus aspects kept in mind and manipulated may be related to memory, or it may be related to the amount of material that can be integrated and kept available at any given period of time.

Crossman and Szafran (1956) also found that age was in different relation to different types of task variations. They found age differences in speed of response greater on the easy than the difficult levels of both a card-sorting and a weight-lifting task. However, with a different card-sorting task, similar in some respects to that of Exp. I in the present study, they found an increase in age difference in sorting speed with the first step of task difficulty, and then a

constant age difference with further difficulty. Their results, especially when viewed in the context of the present and other studies (e.g., Birren & Botwinick, 1955; Clay, 1954; Kay, 1955), suggest that a general theory of aging, as for example, changes in channel capacity, or response reviewing tendencies, may be appropriate only to a specific class of tasks. It may be necessary to categorize tasks into classes by specific characteristics and to develop theories of aging for each class of tasks independently.

The change of task instructions or requirements, examined in Exp. II, indicated that both age groups were statistically similar with respect to the effect of change in task set. Interpretation of this finding however, may need to be qualified because of two reasons. First, change of task set involves modifiability of behavior and in recent studies, age differences in modifiability were found (Botwinick, Brinley, & Robbin, 1958b; Botwinick, Brinley, & Robbin, 1959; Botwinick, Robbin, & Brinley, 1959; Heglin, 1956). It may be, however, that like other descriptions and explanations of aging processes, modifiability of behavior is specific to the class of task or situation. A second reason why interpretation of the results of Exp. II need to be qualified is that ". . . slow people are less likely to suffer from interference phenomenon than are fast people . . ." (Brown, 1914, p. 291). This conclusion was reached by Brown when he compared fast and slow performers with respect to a combination speed and error score of sorting a pattern and then sorting a different one. The implication is that if interference is less with slow performers, the relative slowness of the elderly may have hidden a relative inability to alter the task set. It should be noted that with their slower performances, the elderly did not show less interference effects but instead showed nonstatistically greater interference effects.

## Summary

The effect of age on speed of card-sorting performance was investigated in relation to difficulty of the task (Exp. I), change in task set (Exp. II), and complexity of the task (Exp. III). In addition, the roles o feducation and consecutive sorting trials were studied in relation to the task variables.

In general, it was found that:

1. Card-sorting speed decreased with increased age, task difficulty, and task complexity.

2. Task complexity tended to make for disproportionate slowing of card-sorting performance in the elderly. Task difficulty had the same effect but only when covaried with complexity. This was most apparent in the task that was both most difficult and most complex.

3. Card-sorting speed decreased also with changed task set. This decrease, however, was independent of age.

4. Rate of improvements in speed of card sorting was greatest with the most difficult and complex task.

5. Age differences were not found in rate of improvement or learning. This was true independent of task difficulty, when the task set was changed, and independent of the complexity of the task.

6. The two levels of education were not significantly different in card-sorting speed.

These results were discussed in the context of age differences in perceptual and cognitive functioning.

## References

Anderson, T. W. *An introduction to multivariate statistical analysis.* New York: Wiley, 1958.

Birren, J. E., & Botwinick, J. Speed of response as a function of perceptual difficulty and age. *J. Geront.*, 1955, **10**, 433–436.

Botwinick, J., Brinley, J. F., & Robbin, J. S. The interaction effects of perceptual difficulty and stimulus exposure time on age differences in speed and accuracy of response. *Gerontologia*, 1958, **2**, 1–10. (a)

Botwinick, J., Brinley, J. F., & Robbin, J. S. Task alternation time in relation to problem difficulty and age. *J. Geront.*, 1958, **13**, 414–417. (b)

Botwinick, J., Brinley, J. F., & Robbin, J. S. Modulation of speed of response with age. *J. genet. Psychol.*, 1959, **95**, 137–144.

Botwinick, J., Robbin, J. S., & Brinley, J. F. Reorganization of perceptions with age. *J. Geront.*, 1959, **14**, 85–88.

BROWN, W. Habit interference in sorting cards. *Univer. Calif. Public. Psychol.*, 1914, 1, No. 4, Pp. 269–321.

CLAY, H. M. Changes of performance with age on similar tasks of varying complexity. *Brit. J. Psychol.*, 1954, **45**, 7–13.

CROSSMAN, E. R. F. W., & SZAFRAN, J. Changes with age in the speed of information-intake and discrimination. *Experientia Supplementum IV.* Symposium on Experimental Gerontology. Basel: Birkhauser, 1956. Pp. 128–135.

HEGLIN, H. J. Problem solving set in different age groups. *J. Geront.*, 1956, 11, 310–317.

KAY, H. Some experiments on adult learning. In *old age in the modern world.* Edinburgh and London: E. & S. Livingstone, 1955, Pp. 259–267.

WELFORD, A. T. *Ageing and human skill.* London: Oxford Univer. Press, 1958.

(Received January 23, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 1, 1960

# ONSET VERSUS TERMINATION OF A STIMULUS AS THE CS IN EYELID CONDITIONING[1]

WESLEY J. HANSCHE[2] AND DAVID A. GRANT

*University of Wisconsin*

The purpose of the investigation was to determine whether the termination of a visual stimulus had the same effect as the onset in serving as a CS in eyelid conditioning.

The interval between the onset of the CS and the onset of the UCS (CS-UCS interval) has been shown to be critical in conditioning in a number of researches using various trace and delayed CSs and various CRs. Usually a CS-UCS interval of about .5 sec. is optimal (Bernstein, 1934; Kimble, 1947; McAllister: 1953a, 1953b; White & Schlosberg, 1952; Wolfle: 1930, 1932). Indirect conditioning of the pupillary response may be an exception to this general rule (Gerall & Woodward, 1958). The dependency of conditioning upon the interstimulus interval may be used to investigate the relative efficacy as a CS of the termination and onset of a stimulus, because long interstimulus intervals result in very little conditioning or actual extinction (McAllister, 1953b), and unusual latency distributions (Boneau, 1958). In this investigation, therefore, a stimulus of 1.5 sec. duration was used, with the UCS given either .15, .35, .55, or .75 sec. after the onset of the CS and at .15, .35, .55, and .75 after the termination of the CS in eight different groups.[3] If termination of the stimulus is ineffective as a CS, the latter four groups should show less conditioning than the groups with .15, .35, .55, and .75 sec. between onset of CS and onset of UCS. Conversely, if essentially the same function is found relating rate of conditioning to the interval between CS and UCS onset as is found between CS termination and UCS onset, then the argument favors the proposition that onset and termination of a stimulus are equally effective as the CS.

[1] This research was supported partly by the Research Committee of the Graduate School of the University of Wisconsin with funds granted by the Wisconsin Alumni Research Foundation and partly by a grant from the National Science Foundation to the junior author.

[2] Now at Tulane University.

## METHOD

*Apparatus.*—The apparatus was the same as that used in other experiments from the University of Wisconsin laboratories (Grant, Schipper, & Ross, 1952; Prokasy, Grant, & Myers, 1958). The movement of the eyelid was picked up by a microtorque potentiometer,[4] amplified, and differentiated, so as to give an ink recording of the eyelid responses.

The CS consisted of a light in a 10-cm. circular milkglass window which lasted 1.5 sec. with an illuminance of about 160 mL. The UCS was a corneal air puff of .01 to .02 sec. duration and of sufficient strength to elicit a sharp uniform eyelid reflex. The *Ss* were run under standard light-adapted conditions in order to minimize sensitization of the

[3] Subsequent calibration shows that .04 sec. should be added to all CS-UCS intervals and latencies reported in this experiment.

[4] The use of the microtorque potentiometer which puts virtually no load on the eyelid was originally suggested to the authors by K. W. Spence of the State University of Iowa, but we modified the Iowa arrangements by coupling the shaft of the potentiometer to the eyelid by means of Scotch Tape attached to a length of piano wire which passes into a length of hypodermic tubing, and the tubing is fastened to the potentiometer shaft by means of a hinged coupling.

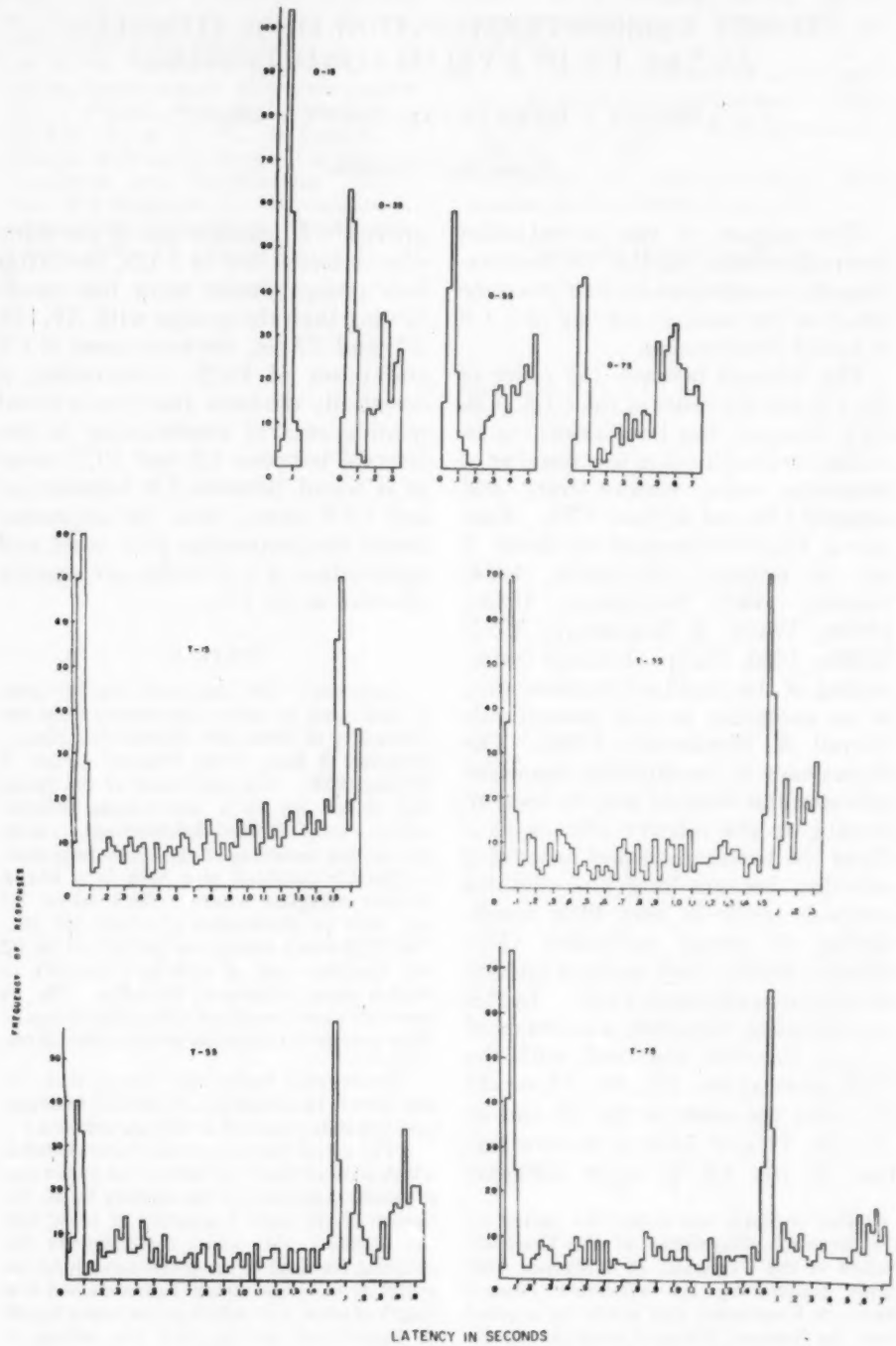


FIG. 1. Latency distributions of all eyelid responses from the four onset groups (above) and the four termination groups below.

beta response (Grant & Norris, 1947). Except in the special groups mentioned below, the corneal air puff was applied to the left eye, and the movements of the right eyelid were recorded.

*Procedure.*—All *Ss* were run for one day only, receiving 40 acquisition trials with 100% reinforcement immediately followed by 10 extinction trials. The intertrial interval varied between 15 and 45 sec. with a mean of about 30 sec.

*Experimental design.*—The primary experiment was a 2 × 4 orthogonal design, with the four interstimulus intervals of .15, .35, .55, and .75 sec. Four groups (O groups) were run with CS-UCS intervals of .15, .35, .55, and .75 sec. from *onset* of the light stimulus to the onset of the UCS. The second set of four groups (T groups) used the same intervals, but they were measured from the *termination* of the light stimulus. Ten *Ss* were randomly assigned to each group by successive replication. Later eight more groups of five *Ss* each were run using a very short light stimulus (recorded as .05 sec. but actually shorter due to the latency of the filament of the 60-w. bulb). The CS-UCS interval, measured from actual onset of CS to onset of UCS was .15, .35, .55, and .75 sec., and for each CS-UCS interval one group of *Ss* had the UCS delivered to the right eye, and another group had the left UCS delivered to the eye.

*Subjects.*—The *Ss* were 120 men and women students from introductory psychology classes at the University of Wisconsin. They were instructed to maintain a passive attitude toward the experiment and not to influence their natural reactions to the stimuli.

## Results

Orderly data were obtained showing that the termination of the stimulus was as effective as its onset in serving as a CS in eyelid conditioning, and, further, that both onset and termination CS-UCS intervals had the same functional relationship to rate of conditioning. As secondary findings, some striking new data on alpha conditioning were obtained, and contralateral conditioning was found to be almost the same as ipsilateral conditioning.

*Acquisition.*—Latency distributions are presented for all eight groups in the primary experiment in Fig. 1. Here the number of eyelid responses is plotted against latency for all responses during acquisition. In the four onset groups (O groups) the left-hand mode of responses consists of primary reflexes or alpha responses to light and the right-hand mode consists of CRs. For the termination groups (T groups) the left-hand mode consists of alpha responses, the next mode to the right consists of off effect responses (ORs), and the responses further to the right consist of CRs. Occurrence of stimulus termination evidently prevents development of

TABLE 1

Analyses of Variance of Total Frequency of Alpha Responses, Off Responses, and CRs During Acquisition, Subjected to the Arcsin Transformation

| Source | Alpha Responses | | Off Responses | | CRs | |
|---|---|---|---|---|---|---|
| | *df* | *F* | *df* | *F* | *df* | *F* |
| Onset-Termination (OT) | 1 | 0.10 | | | 1 | 0.13 |
| CS-UCS interval (I) | (3) | | (3) | | (3) | |
| Linear | 1 | 5.33* | 1 | 3.38 | 1 | 64.75** |
| Quadratic | 1 | 1.35 | 1 | 3.07 | 1 | 9.66** |
| Cubic | 1 | 0.10 | 1 | 0.44 | 1 | 0.98 |
| OT × I | 3 | 0.46 | | | 3 | 2.38 |
| Error (*MS*) | 74 | (251.7) | 36 | (229.6) | 72 | (220.5) |

* $P = .05$.
** $P = .01$.

long latency CRs as reported by Boneau (1958). The OR is apparently analogous to the alpha reflex to onset of light. It will be noted that the number of alpha responses declined steadily with increasing CS-UCS interval in the onset groups, showing that the level of alpha conditioning (Hilgard & Biel, 1937; Weber & Wendt, 1942; Grant & Adams, 1944) is related to the CS-UCS interval. This is a new observation. The termination groups show some evidence of this tendency in alpha responding, although the T − .75 group departs from the general trend.

Table 1 presents a summary of the analysis of variance (Grant, 1956) of frequency of responses during acquisition. There is a significant linear trend in the alpha responses as a function of CS-UCS interval. There is no significant interaction between CS-UCS interval and onset versus termination.

Although the alpha responses and ORs seem dependent upon the CS-UCS interval, their primary tendency during the experiment was to adapt out. This is shown in Fig. 2 where both the alpha responses and the ORs decrease to about half the initial frequency during the course of the experiment. There is no evidence of an acquisition trend, so that alpha conditioning is represented primarily by a resistance to adaptation rather than as an actual increase in frequency of responding (cf. Grant & Adams, 1944).

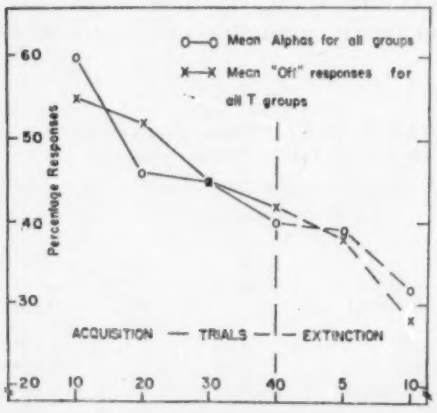


FIG. 2. Percentage frequency of alpha responses and off-responses (ORs) plotted in 10-trial blocks during acquisition and 5-trial blocks during extinction.

The OR follows the termination of the CS at approximately the same latency as that between the onset of the CS and the alpha response. There is no indication that the OR is dependent upon the interval between termination of the CS and onset of the UCS. In Table 1 there are no significant trend components in the frequency of ORs as a function of the CS-UCS interval.

The acquisition and extinction of the CRs in the eight groups are shown in Fig. 3, where percentage CRs is plotted against trials in 10-trial blocks during acquisition and 5-trial blocks during extinction. The usual acquisition trends are shown except in the short CS-UCS interval groups. Progressively more conditioning is found with the higher CS-UCS intervals both with respect to onset and termination, and the linear and quadratic components of this relationship to CS-UCS interval are significant as shown in Table 1. There is no significant difference in the total frequency of CRs between onset and to termination groups.

*The CS-UCS interval effect.*—Plotting rate of conditioning against the CS-UCS interval during acquisition presents certain technical difficulties. For the short CS-UCS interval the occurrence of the UCS and reflexes to the UCS may hide the CRs so that the level of conditioning cannot be determined directly from the records. In addition, the longer the CS-UCS

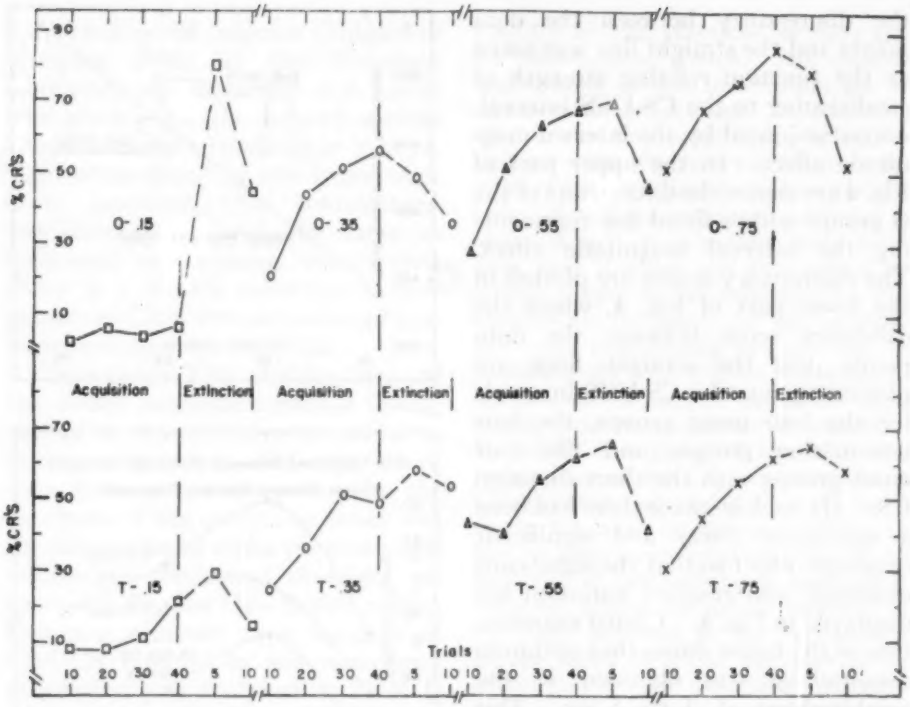


FIG. 3. Percentage CRs plotted in 10-trial blocks during acquisition and 5-trial blocks during acquisition and 5-trial blocks during extinction for the four onset groups (O groups) and the four termination groups (T groups).

interval the greater the opportunity of a random blink occurring during the anticipatory interval before the UCS occurs. Even if increased conditioning did not result in a higher probability of a response occurring during any part of the CS-UCS interval, the longer the CS-UCS interval available for inspection the more responses will be observed. This interval magnitude effect can be handled in various ways. For example, Kimble (1947) inserted test trials on which the UCS was not given in order to get a clear view of all of the CRs. McAllister (1953a) used the first block of acquisition trials to calculate the random rate or operant level of responding during the CS-UCS interval for each *S* and subtracted this amount from the frequency of CRs on the following blocks of trials. In the current experiment a statistical method was employed which may have the advantage of utilizing the data more fully. As an approximation it was assumed that the longer the interval before the UCS the more likely a response was to occur owing simply to the length of the CS-UCS interval and not to the strength of conditioning per se. This implied that over and above general rate of conditioning there would be a linear increase in number of recorded CRs as the CS-UCS interval is increased. Therefore the number of CRs was plotted against the CS-UCS interval and a straight line was fitted by least squares to the observed data points. Assuming the straight line to be primarily interval magnitude effect,

the discrepancy between the data points and the straight line was taken as the function relating strength of conditioning to the CS-UCS interval, uncontaminated by the interval magnitude effect. In the upper part of Fig. 4 are shown the data points of the O groups with a fitted line representing the interval magnitude effect. The discrepancy scores are plotted in the lower part of Fig. 4, where the difference score between the data points and the straight lines are plotted against the CS-UCS intervals for the four onset groups, the four termination groups, and the four onset groups with the short duration CS. (In each instance there had been a significant linear and significant quadratic effect so that the significant quadratic and residual variation are displayed in Fig. 4. Casual examination of the figure shows that optimum conditioning was obtained in the neighborhood of .3 to .5 sec. This may be somewhat shorter than the optimum interval that has always been found for the CS-UCS interval function by other investigators in eyelid conditioning using other techniques of correcting for the CS-UCS interval magnitude artifact.)

The most important observation from the standpoint of this experiment is that the termination function in Fig. 4 is essentially the same as the onset function. Furthermore the duration of the CS per se seems to make little difference because the short duration CS produces the same function as the long duration so that trace and delayed functions are essentially identical. This observation is confirmed by the fact that there is no significant interaction between the three groups and the quadratic components of these trends.

*Contralateral vs. ipsilateral conditioning.*—Although no analysis will be presented it may be remarked that the acquisition trends and the CS-UCS interval trends were identical for those *Ss* whose responses were recorded from the puffed right eye and those whose left eye was puffed and right eye responses were recorded. There was a 5–10% reduction in frequency on the contralateral eyelid. Although the primates generally give

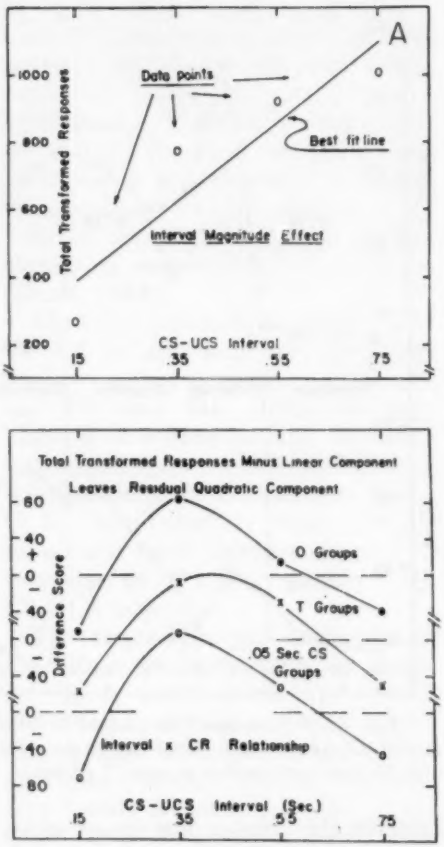


FIG. 4. Determination of the effect of CS-UCS interval on frequency of eyelid CRs. In the upper panel a least squares line representing estimated interval magnitude effect is fitted to the transformed (arcsin) total frequency of CRs in the four onset groups. In the lower panel the discrepancies from the fitted line are plotted along with similar discrepancies from the four termination groups and four short-duration CS groups.

a bilateral eyelid response (Hilgard & Marquis, 1936), in the Wisconsin laboratory an occasional *S* is found who gives a greatly reduced contralateral response to puffing the left eye. The results found in this experiment show, however, that contralateral conditioning is about the same as ipsilateral conditioning, except that there is a 5–10% reduction in frequency of CR if the contralateral eyelid movements are recorded.

*Extinction.*—The distributions of the eyelid response latencies during extinction showed the same modes as during acquisition, although there were fewer responses owing to the shortness of the extinction series and the adaptation of alpha responses and ORs. The frequency of alpha responses, CRs, and ORs during extinction was analyzed as a function of CS-UCS interval in the same way acquisition was analyzed in Table 1. Alpha responses showed the same significant linear decrease with size of CS-UCS interval as was shown in acquisition. Frequency of ORs was not affected by the CS-UCS interval just as in acquisition. In the case of CRs, however, the extinction data showed no significant trend components in contrast to the distinct linear and quadratic trends during acquisition. This was due at least in part to the brevity of the extinction series but also to an "aberrant" result in the extinction of the 0–.15 group. (See Fig. 3.) This aberrant result was not found with the corresponding group given a short duration CS.

## Discussion

The primary finding of the experiment was that the termination of the stimulus may serve almost exactly the same function as the onset of the same stimulus in the occurrence of conditioning. This is reflected in the essential identity of the CS-UCS interval functions for onset and for termination groups. The general implications of this observation are of considerable importance. It has long been known that conditioning is usually optimal when the interval between the onset of the conditioned stimulus and the onset of the unconditioned stimulus is about .5 sec. It is evident from these data that if there are two salient or critical points to the stimulus—in this experiment, the onset and the termination—the conditioning process will be optimal if the UCS is given in the neighborhood of .5 sec. of *either* of these critical features, onset or termination. If the UCS is given .5 sec. after the termination of the CS, the onset of the CS appears to have little or no relevance to the formation of the eyelid CR. In fact a very careful examination of all responses at various latency ranges that occurred during the actual presence of the light stimulus in the termination groups failed to reveal any sign of an acquisition or an extinction tendency among the responses in any latency range. They were, evidently, random blinks. At least they did not show any of the usual tendencies to increase with reinforcement and decrease with extinction.

There is some indication that the short CS-UCS interval, although it does not produce much conditioning in the ordinary sense, does produce a fairly high level of alpha conditioning or reflex sensitization. This was shown during both acquisition and extinction in the onset groups. In spite of the alpha conditioning, the predominant trend of the alpha responses was downward during the experiment, indicating that reflex adaptation had a greater effect than reflex sensitization due to the presence of the corneal puff.

The OR was found to occur with about the same frequency and latency, relative to the termination of the CS, as did the alpha response to the onset of the CS. Beside the almost identical latency, frequency, and adaptation trends of the alphas and ORs there was a correlation of .72 between the frequency of alphas

and frequency of ORs for the 40 *Ss* for whom both responses were available.

## Summary

An experiment with 120 human *Ss* was performed to compare the effectiveness of stimulus termination with stimulus onset as the CS in eyelid conditioning.

The principal results were as follows:

1. The effect of the CS-UCS interval on the level of conditioning during acquisition trials was exactly the same whether the onset or the termination of the light served as the CS. Optimal conditioning occurred with a CS-UCS interval of approximately .5 sec. or less.

2. In addition to the usual primary reflex or alpha response to light, an off response (OR) was observed which was closely analogous to the alpha response except that it occurred as a response to termination of the light. The ORs were practically identical with the alpha responses in frequency, latency, and adaptation trend.

3. There was evidence of alpha conditioning that was most marked with the short CS-UCS intervals, but the corresponding trend with ORs was not significant.

It was concluded that the termination of a stimulus is just as critical a feature for conditioning as its onset. If the termination of the stimulus occurs .3–.5 sec. before the UCS, conditioning will take place to termination rather than onset which indicates a certain interesting flexibility in the learning mechanism.

## References

Bernstein, A. L. Temporal factors in the formation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–197.

Boneau, C. A. The interstimulus interval and the latency of the conditioned eyelid response. *J. exp. Psychol.*, 1958, **56**, 464–471.

Gerall, A. A., & Woodward, J. K. Conditioning of the human pupillary dilation response as a function of the CS-UCS interval. *J. exp. Psychol.*, 1958, **55**, 501–507.

Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Grant, D. A., & Adams, J. K. "Alpha" conditioning in the eyelid. *J. exp. Psychol.*, 1944, **34**, 136–141.

Grant, D. A., & Norris, E. B. Eyelid conditioning as influenced by the presence of sensitized Beta-responses. *J. exp. Psychol.*, 1947, **37**, 423–433.

Grant, D. A., Schipper, L. M., & Ross, B. M. The effect of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44**, 203–210.

Hilgard, E. R., & Biel, W. C. Reflex sensitization and conditioning of eyelid responses at intervals near simultaneity. *J. gen. Psychol.*, 1937, **16**, 223–234.

Hilgard, E. R., & Marquis, D. G. Conditioned eyelid responses in monkeys, with a comparison of dog, monkey, and man. *Psychol. Monog.*, 1936, **47**(2), (Whole No. 212).

Kimble, A. Conditioning as a function of the time between conditioned and unconditioned stimuli. *J. exp. Psychol.*, 1947, **37**, 1–15.

McAllister, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 417–422. (a)

McAllister, W. R. The effect of eyelid conditioning of shifting the CS-UCS interval. *J. exp. Psychol.*, 1953, **45**, 423–428. (b)

Prokasy, W. F., Jr., Grant, D. A., & Myers, N. A. Eyelid conditioning as a function of unconditioned stimulus intensity and intertrial interval. *J. exp. Psychol.*, 1958, **55**, 242–246.

Weber, H., & Wendt, G. R. Conditioning of eyelid closure with various conditions of reinforcement. *J. exp. Psychol.*, 1942, **30**, 114–124.

White, C. T., & Schlosberg, H. Degrees of conditioning of the GSR as a function of the period of delay. *J. exp. Psychol.*, 1952, **43**, 357–362.

Wolfle, H. M. Time factors in conditioning finger withdrawal. *J. gen. Psychol.*, 1930, **4**, 372–379.

Wolfle, H. M. Conditioning as a function of the interval between the conditioned and original stimulus. *J. gen. Psychol.*, 1932, **7**, 80–103.

(Received January 26, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 1, 1960

# THE SOMATIC BACKGROUND OF ROTE LEARNING[1]

R. N. BERRY AND R. C. DAVIS

*Indiana University*

A previous report (Berry & Davis, 1958) in this series discussed muscle activity in relation to rote learning of nonsense syllables. The present report describes a further experiment on rote learning, with some procedural changes, in which the striate responses of the jaw and arm are re-examined, and heart rate, electrocardiogram, volume pulse, and GSR measures are also evaluated.

As in the previous report, the primary interest was in establishing empirical relations rather than in searching for support for a theoretical position.

## METHOD

*Subjects and apparatus.*—The *S*s were male students, paid volunteers from the summer student population. The cooperation of the *S*s was excellent, and there seems no reason to question their motivation.

Only 16 *S*s are considered in the treatment of the data; there were a number of others who have been excluded because of failure to follow instructions, failure to reach criterion within 25 trials, or malfunction of the apparatus. The 25-trial limitation was set early in experimentation because of the prohibitive cost of operating the recording apparatus for greater numbers of trials.

Each *S* was placed in a chair in the experimental room; the room contained a memory drum and was separated from a recording room by a sound-reducing partition. Bipolar surface electrodes, with saline paste, were fastened over the jaw (masseter) and right forearm (extensor) muscles of *S*. The EKG and the heart rate were obtained from electrodes in the Lead I position, and the volume pulse was obtained with an electric finger plethysmograph. The palmar GSR was recorded from yet another set of leads. The details of the recording apparatus are described in a previous report (Ellson, Davis, Burke, & Saltzman, 1952).

*Procedure.*—The memory drum was so arranged that *S* was presented with a blank white space in the memory drum for the first 2 sec. This was the interval in which *S* was instructed to make his verbal anticipation of the upcoming item. A light then flashed momentarily in the memory drum aperture to indicate that the anticipation interval was terminated. Two seconds later the item on the drum appeared and remained in view for 2 sec. The 6-sec. cycle was repeated until all of the items had been presented. The next trial was begun without interruption.

The nine items on the memory drum were two-place numbers presented to each *S* in the same order: 35, 29, 26, 32, 39, 27, 24, 38, 31. In the instructions, preceding the learning session, *S*s were told that there was a list of numbers to be memorized and that the numbers might range from 21 through 39. They were also told that there were no duplications within the list. Further instructions were to the effect that the memory drum would continue to rotate until every item in the list was correctly anticipated within the same trial and that *S* was to anticipate each upcoming number, including those on the first trial, even if the anticipations were only guesses. This last instruction was intended to eliminate the difficulty of interpreting the data in those cases where no understandable anticipation was made, a problem which had arisen in the previous experiment.

*Data analysis.*—The kymograph record contained tracings from all the leads, as well as a tracing indicating the action of the memory drum. Thus the records for an entire learning session could be divided into 1-sec. intervals and the appropriate measurements made for each variable for each 1-sec. interval. The measures are analyzed in terms of the basic 6-sec. cycle of the memory drum, covering the anticipation, delay, and information intervals. Further analyses are made in terms of correct and incorrect anticipations and in relation to speed of learning. The latter is defined by the number of trials to a criterion of one correct trial, and four learning speeds are differentiated: 6–10 trials,

[1] This research was carried out under Contract Nonr 908(03). R. C. Davis is the principal investigator under this contract. The senior author is indebted to the Graduate School of Indiana University for financial assistance in carrying out this study.

11–15 trials, 16–20 trials, and 21–25 trials. There are 6, 5, 2, and 3 *S*s in these respective groups, and they are identified hereafter as Groups I, II, III, and IV.

## Results

*The MAP from the masseter.*—The recording device integrated the output voltage of the masseter for .1-sec. intervals. Thus for any 1 sec. of the kymograph record, there were 10 values available. In all cases, only the largest of these 10 values has been considered in the data treatment. This practice is in line with previous studies and assumes that it is the maximum response in which the primary interest lies.

Six MAP values (representing each second of the 6-sec. memory drum cycle) for the correct anticipations and six values for the incorrect anticipations were obtained for each *S*. This was done by first obtaining medians across serial positions for each trial and then medians across trials.

The *S*s were grouped according to learning speeds as indicated above, and the mean values, second by second, for each group for correct and incorrect responses are presented in Fig. 1. A variety of hypotheses might be tested, but most importantly those that were supported in the earlier study.

It is quite apparent, in the present experiment, that there is a tendency for the over-all activity to decrease as learning speed decreases; the poorest 5 *S*s (Groups III and IV) show very little activity throughout. The same trend, with a reversal for the very poorest *S*s, existed in the earlier data, and a significant relation between speed of learning and an over-all measure of jaw-forehead activity was also found. In the present experiment, only jaw data are available. However, for the over-all mean for

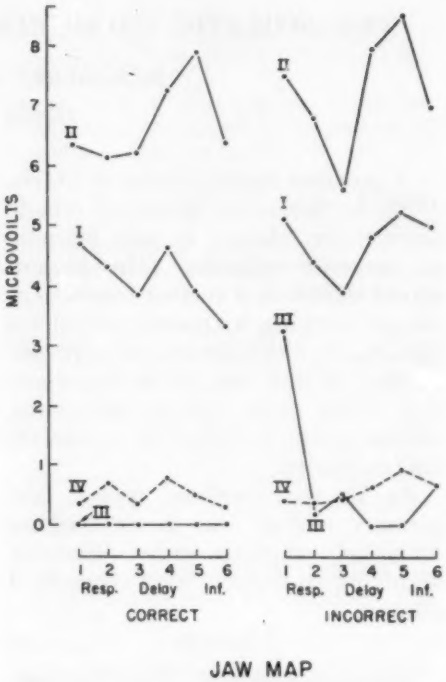


FIG. 1. The mean MAP values for each second of the 6-sec. memory drum cycle for correct and for incorrect anticipations. The roman numerals identify the learning groups, I being the fastest and IV the slowest.

each *S*, the Kruskal-Wallis test yields $.06 > P > .05$. This is very close to the value obtained in the previous experiment, and, correcting for the small *N*s in some of the groups, is significant at the .05 level. That there are no high voltages for the slowest learners in this study, as opposed to the earlier study, might be attributed to the truncated distribution produced by considering only those *S*s who reached criterion within 25 trials.

It is possible that the significant correlation between learning speed and general level of tension in the masseter is a function of temporal changes over trials. However, this would seem an unlikely explanation since the over-all tension levels for the correct and in-

correct responses appear to be very similar within learning groups. If temporal factors were significant, the incorrect responses, tending to come earlier in the series, would show a high over-all level since the rapid learners showed a higher over-all level.

In the earlier study there was a significant nonmonotonic relation between speed of learning and the difference between the changes from response to information interval for correct and for incorrect responses. In the present experiment this trend is confirmed, but it is not significant at the .05 level. The small difference in results between the two experiments might easily be attributed to procedural differences, making a clear-cut comparison difficult.

*The MAP from the arm extensor.*—The data can be discussed very briefly. As in the previous experiment in which the arm MAP was recorded during rote learning, no systematic relation between the muscle action measure and any learning measure could be demonstrated.

*The electrocardiogram.*—The measure of the EKG was the height of the line from the Q to the R points, and this height was converted to a percentage of the mean rest height. The mean rest height was the mean of the values for seven consecutive heart beats occurring about 30 sec. before the learning session was begun. Because the measure is relative, it is not appropriate to make any interpretation which assumes absolute values.

Figure 2 shows the mean second by second values across all serial positions and all trials for each of four learning groups, with a further breakdown into correct and incorrect responses. There is a tendency for the information interval to show large values, and this is quite striking in the case of the incorrect responses. Since these trends occur for all learning groups, the data for all *S*s are combined and presented in Fig. 3. The most striking feature in this plot is the difference between the correct and the incorrect response magnitudes in the information interval. If we apply a simple binomial sign test, the mean of Sec. 5 and 6 values is significantly larger ($P < .01$) for the incorrect responses than for the correct responses. (There is a tendency, although not significant at the .05 level, for the size of this difference to be larger for the rapid learners.) A more rigorous argument concerning the significant finding is made by a further manipulation of the data. For each *S* the mean value of Sec. 5 was subtracted from the mean value of Sec. 4 for the correct responses and for the incorrect responses. The binomial sign test shows a significant difference ($P < .05$) between the difference measures for the correct and for the incorrect responses.

From what is known of the relation between respiratory cycle and the

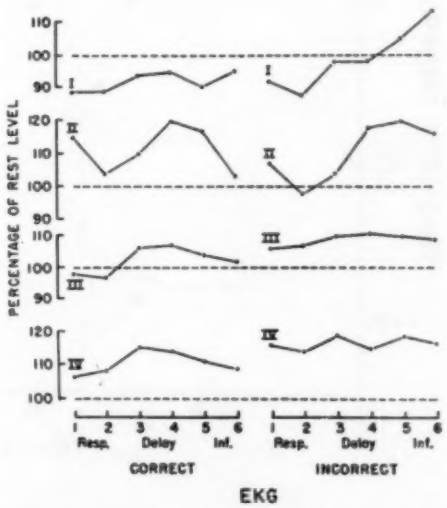


FIG. 2. The electrocardiogram values for each of the four learning groups.

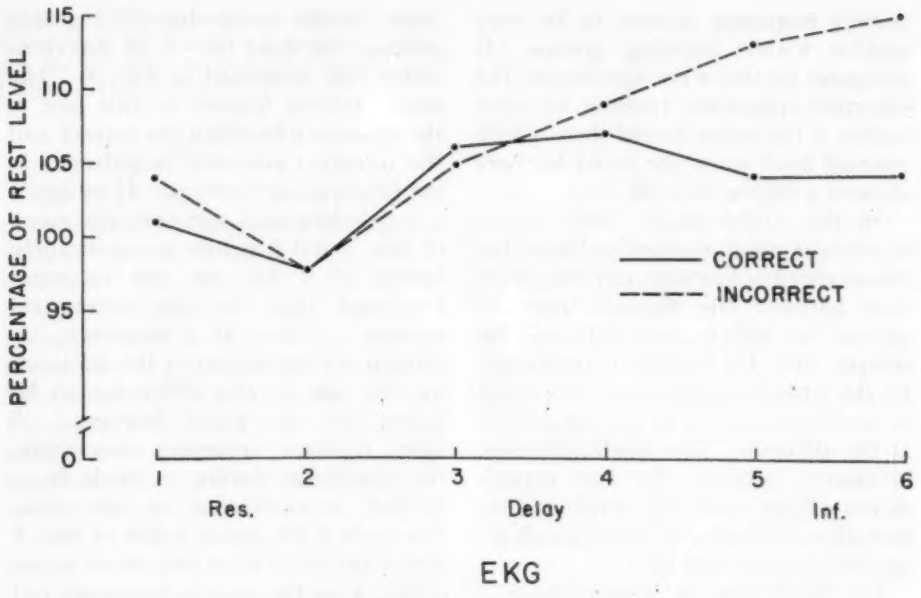


FIG. 3. The data of Fig. 2 combined for all learning groups.

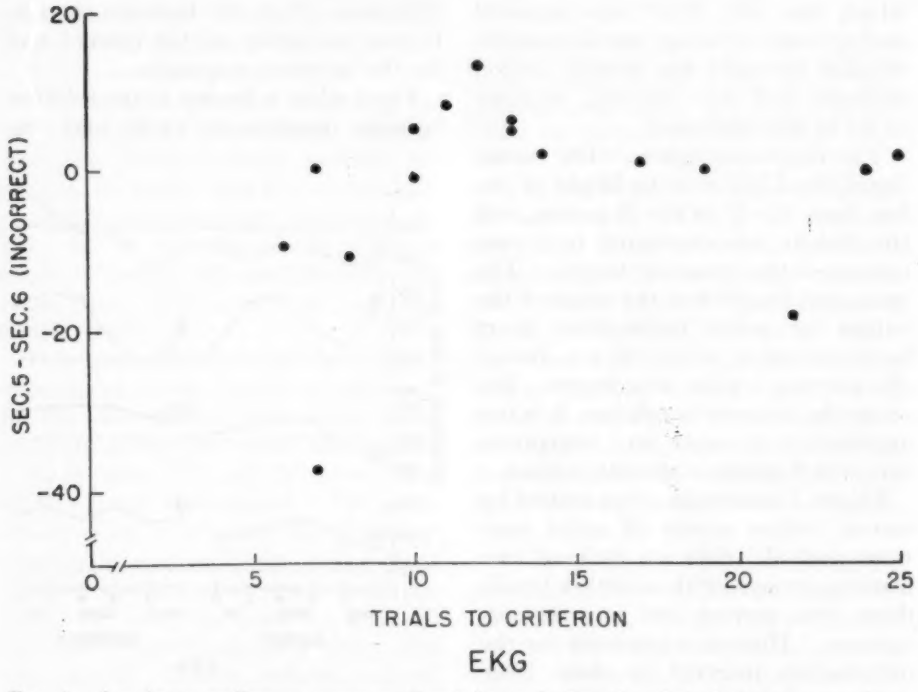


FIG. 4. An electrocardiogram measure (involving only Sec. 5 and 6 of the memory drum cycle) related to speed of learning.

electrocardiogram, it is safe to assume that the larger measures represent an expiratory movement in the breathing pattern. The data would sensibly be interpreted as corrective responses in the case of an incorrect anticipation, the corrective responses otherwise not being required. It may be noted in Fig. 3 that the differences between the values of the rights and wrongs in the response interval are negligible. This gives further emphasis to the importance of the information interval in the learning pattern.

Since the information interval for the incorrect responses appeared to be of interest, one further analysis was made utilizing the Sec. 5 and 6 values. The data for the analysis are summarized in Fig. 4. For the incorrect responses, the mean value of Sec. 6 is subtracted from the mean value of Sec. 5 for each *S* and plotted against the number of trials to criterion. It is apparent that the relation is curvilinear; the Kruskal-Wallis test shows that the relation is significant at the .05 level. Since there is no such systematic relation for the correct anticipations, the nature of somatic responses in the information interval following an incorrect response appears to be critical.

*The volume pulse.*—The magnitude of the volume pulse was measured for the first pulse in each 1-sec. interval and converted to a percentage of rest level. The latter was a mean magnitude for six consecutive seconds in a period about 30 sec. before the memory drum was started.

The mean data for the correct and incorrect responses are presented in Fig. 5. Here again, *S*s are divided into four learning groups, the same ones as used above. Although there

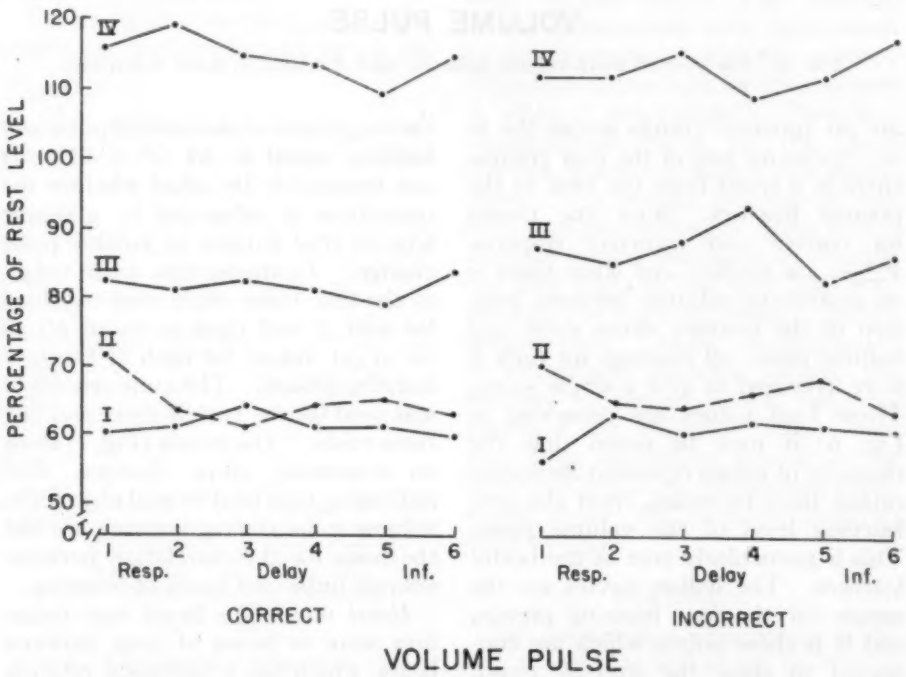


FIG. 5. The mean volume pulse (percentage of rest level) for each of the four learning groups.

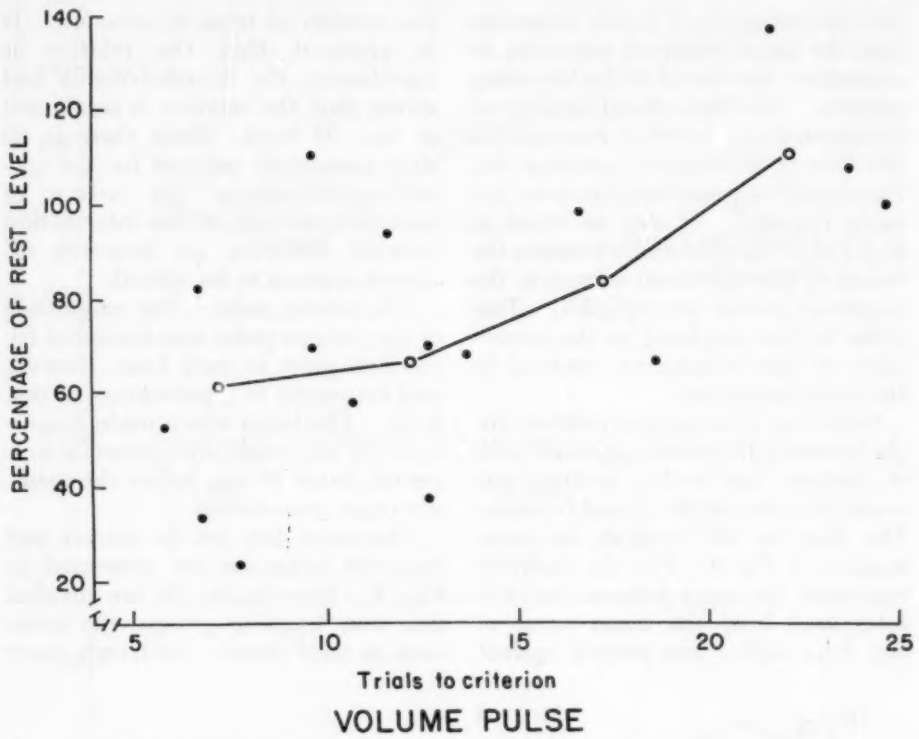


FIG. 6. The over-all mean volume pulse for each *S* related to speed of learning.

are no apparent trends across the 6-sec. cycles for any of the four groups, there is a trend from the best to the poorest learners. Since the trends for correct and incorrect response values are similar, and since there is no systematic relation between position of the memory drum cycle and volume pulse, all readings for each *S* were averaged to give a single value. These final values are presented in Fig. 6; it may be noted that the majority of values represent decreases, rather than increases, from the pre-learning level of the volume pulse. This is particularly true of the better learners. The hollow circles are the means for the four learning groups, and it is these points which are connected to show the over-all trend. The rank-order correlation between the magnitude of the volume pulse and learning speed is .64 ($P < .01$). It can reasonably be asked whether the correlation is influenced by a simple trial to trial pattern of volume pulse change. To answer this, mean values of the first three trials were obtained for each *S* and then averaged across *S*s to get values for each of the four learning groups. The same procedure was used for the middle three and last three trials. The trends (Fig. 7) show no systematic slope changes, thus indicating that trial to trial changes in volume pulse during learning are not the basis for the correlation between volume pulse and speed of learning.

*Heart rate.*—The heart rate measures were in terms of time between beats, which has a reciprocal relation to the usual measure of beats per

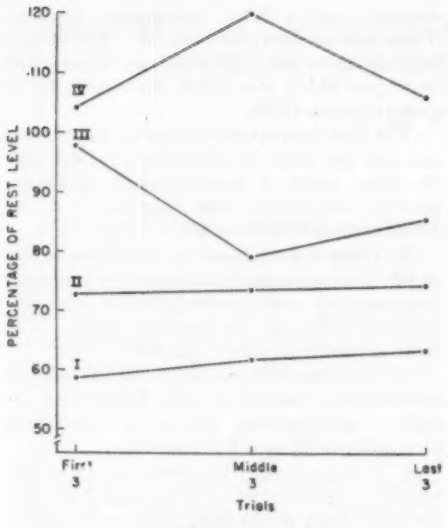


FIG. 7. A mean volume pulse measure for each learning group related to trial position.

minute. Thus the larger values reflect slower heart rates.

The mean data for the four learning groups for both right and wrong responses are presented in Fig. 8. Grouping all *S*s, there is a significant trend for the heart rate to decrease from Sec. 1 to Sec. 5 of the cycle, for both correct and incorrect responses. However, the magnitude of the decrease is not significantly related to speed of learning for either response category, nor are there significant differences in the trends for correct and incorrect responses. Lastly, using a mean heart rate involving all the values for each *S*, no significant correlation between these means and speed of learning could be obtained.

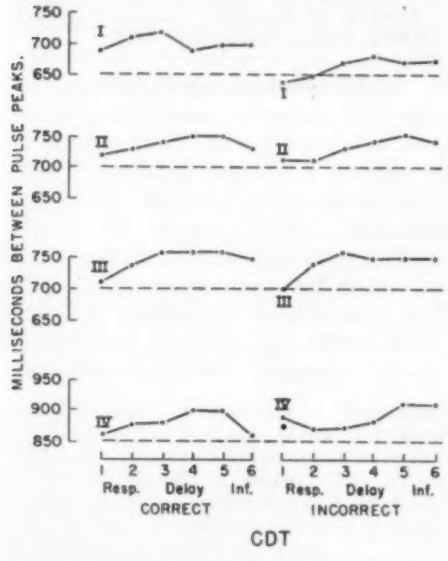


FIG. 8. The heart rate measure for each of the four learning groups.

Heart rate changes are probably a secondary effect of an increased EKG. Stronger heart action would slow pulse rate through the action of the carotid sinus reflex. As a secondary effect, the heart rate changes would have low or negligible correlation with other changes.

*GSR.*—The GSR records were measured in two ways. The absolute levels of resistance were determined for the moment at which the memory drum was started and for the moment at which it was stopped. There is a nonsignificant trend for the rapid learners to exhibit higher skin resistances. The second measure of the GSR was the number of times that a decrease in skin resistance occurred in the first trial, in the trial half way through the learning session, and in the last trial. Again, no significant or systematic relations among the learning speeds or among the trial positions are evident. There were very large differences among *S*s for both GSR measures, differences which would make the negative findings quite tentative.

## DISCUSSION

The physiological adjustments to the task situation as well as the moment to moment changes have been shown to

relate to measures of learning performance. As for the adjustment to the task situation, speed of learning correlates with the magnitudes of the jaw MAP and of the volume pulse. Generally, it seems that the most rapid learners generate a high striate activity in the head region.

The moment to moment changes associated with speed of learning are reflected in the jaw MAP. There is a tendency for the MAP pattern differences between the correct and incorrect anticipations to vary systematically with speed of learning. The same tendency occurs in the EKG data. And the EKG exhibits a significant correlation between temporal patterns for the incorrect responses and speed of learning.

As an over-all generalization, it is contended that the rate of verbal learning is associated with increased physiological activity in the head region and further that this activity is most important in the interval when a correcting response is made following an incorrect anticipation.

## Summary

Sixteen *Ss* learned under instructions to verbally anticipate every item in a list of nine two-place numbers; the learning criterion was one correct trial. The memory drum presented a basic 6-sec. cycle for each number: a 2-sec. anticipation interval, a 2-sec. delay interval, and a 2-sec. information interval. There was no intertrial interval. The following variables were recorded on kymograph tape: jaw MAP, arm MAP, heart rate, EKG, volume pulse, GSR.

The data support the following statements:

1. In the over-all adjustment to the task situation, speed of learning has a significant positive correlation with magnitude of jaw MAP and with volume pulse.

2. There is a tendency for the difference in MAP (jaw) patterns for correct and incorrect responses to vary systematically with individual differences in learning speed. The same tendency occurs in the EKG data.

3. The temporal pattern of EKG magnitudes during the information interval of incorrect anticipations shows a significant correlation with speed of learning.

4. It is concluded that the rate of verbal learning is associated with increased physiological activity in the head region and further that this activity is most critical in the interval when a correcting response is made following an incorrect anticipation.

## References

Berry, R. N., & Davis, R. C. Muscle responses and their relation to rote learning. *J. exp. Psychol.*, 1958, **55**, 188–194.

Ellson, D. G., Davis, R. C., Burke, C. J., & Saltzman, I. J. *A report of research on detection of deception.* Contract N6onr-18011, Indiana University, 1952.

(Received January 31, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 1, 1960

# "ASSOCIATION" OF NONSENSE SYLLABLES FOLLOWING VARIED LEARNING CONDITIONS[1]

BONNIE WEBB CAMP[2]

*Indiana University*

This study was designed to determine whether some of the assumptions concerning the development of word associations and of verbal behavior in general could be investigated directly in nonsense syllable learning experiments which use the association method as a test of learning. The association method itself consists of presenting *S* with a stimulus, usually a word or syllable, with instructions to respond with the first other word or syllable that occurs to him. Traditionally, these responses are called "associations," and in this paper, the term associations will be used only in reference to responses obtained by this procedure.

Even in this narrow sense, however, associations, particularly word associations, have been the subject of considerable interest in both experimental and clinical settings. In spite of the obviously learned nature of word associations, however, the association method has not been used as a test of learning outside of studies of intraserial learning phenomena (McGeoch & Irion, 1952). Although more recent studies (Jenkins & Russell, 1952), have departed from the classical studies of classification and "emotional disturbance" (Bell, 1948; Woodworth, 1938), attempts to observe directly, through experimental manipulation, the conditions of learning which are functionally related to the production of associations have been notably lacking. It would appear that investigating the acquisition of associations in an experimental setting might help to increase our understanding of the role of antecedent events in the production of word associations and of verbal behavior in general. The present study employing nonsense syllables represents a step in this direction.

Trials with paired-associate nonsense syllables provides, in a learning task, the basic elements which are tested by the association method. Presentation of syllable pairs, however, does not in itself eliminate the possibility that differences in associations could be attributed to differences in *S*'s exposure to the syllables in an experimental situation and, consequently, his familiarity with them. If we define familiarity with the syllables as the frequency of occurrence of the syllables as responses, whether alone or in pairs, familiarity with either member of a pair of syllables becomes a potentially important factor which could have an independent effect on results of an association test.

The following study, therefore, was designed to determine the effect of variation in the number of paired presentations of syllables and variation in familiarity with either member of a pair of syllables on the syllables produced in an association test.

[1] This study is based on a dissertation presented to the Graduate School of Indiana University in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express appreciation to A. M. Buchwald, committee chairman and to D. C. Beier, I. J. Saltzman, and H. E. Driver.

[2] Now at University of Colorado School of Medicine.

## METHOD

The general plan of this study involved visual presentation of nonsense syllables in varying frequencies of occurrence, singly and in pairs. The *Ss* were simply asked to copy the syllables on separate sheets of paper as they were exposed. This insured that each *S* used each syllable, or pair of syllables, as a response the required number of times and kept his method of responding consistent with the method of presentation. After *Ss* were presented with single syllables in a random sequence, followed by pairs of syllables in a random sequence, an association test of learning was introduced.

*Apparatus.*—Twelve nonsense syllables with 40% association value were chosen from Glaze's (1928) list of syllables. Two lists of six syllables each were selected at random within the limits of criteria insuring the least possible amount of repetition of vowel and consonant sounds. Pairs of syllables consisted of one syllable from each list. Syllables in one list were designated stimulus syllables (S syllables). Those in the other list were designated response syllables (R syllables). The syllable pairs were referred to as S-R.

The selected syllables were printed in the center of $1\frac{1}{2} \times 2\frac{1}{2}$-in. cards. The S cards were those bearing S syllables, and R cards those bearing R syllables. On each S-R card, an S syllable and an R syllable were printed side by side with the S syllable always appearing on the left. The same R syllable was always paired with a given S syllable and thus gave a total of six different S-R cards.

The syllables were then arranged in three different types of series. The first type consisted of nine sets of both S and R cards. Each set contained either 2, 6, or 10 occurrences of each S syllable, and either 2, 6, or 10 occurrences of each R syllable arranged in a random order. All *Ss* receiving the same combination of S and R occurrence were presented with the same set of cards in the same order.

The second type of series consisted entirely of the six S-R cards. These cards were ordered randomly in sets of six cards so that each S-R pair occurred once in each set. Cards were arranged in a total of 10 such sets. By presenting either none of these sets, the first 6, or all 10 of these sets, it was possible to present different groups with 0, 6, or 10 exposures of each S-R pair. The nine sets containing both S cards and R cards and the series of S-R cards were combined in such a way that each group of five *Ss* received 2, 6, or 10 exposures of each S syllable, 2, 6, or 10 exposures of each R syllable and 0, 6, or 10 exposures of each S-R pair. This resulted in a total of 27 possible combinations of each condition.

The third type of series consisted of one each of the six different S cards arranged in a random order. These cards were presented on test trials to all *Ss*.

*Subjects.*—The *Ss* were 135 undergraduates enrolled in introductory psychology classes at Indiana University. They were divided at random into 27 groups of five each so that each of the possible combinations described above was administered to five *Ss*.

*Procedure.*—The *Ss* were seen individually in a small room ordinarily used for administering psychological tests. They were seated comfortably at a small table across from *E*. A lower table was placed beneath this table and to one side of *E*. This allowed *E* to conceal from *S* all cards except the one being presented. A stack of $2\frac{3}{4} \times 2\frac{1}{4}$-in. paper and a pencil were placed on the table in front of *S*. The room was otherwise bare.

Each *S* was read the following instructions: "I have a series of cards here with nonsense syllables on them. Do you know what a nonsense syllable is?" (If *S* indicated that he did not know, an explanation was given.) "I am going to show these to you one at a time. Write each one on the top piece of paper, turn the paper over immediately and do not look at it again. All you have to do is write each one down." The *Ss* were then shown one of the sets of cards containing both S and R cards. The *E* presented the cards manually, one at a time. Each card was removed and replaced by another when *S* had finished copying the syllable.

After completion of the first series, paper on which *S* had written was removed. Those *Ss* receiving a series of S-R pairs were read the following instructions: "I have another series of cards but this time there are two nonsense syllables on each card. Write both of them on the top piece of paper, turn the paper over immediately and do not look at it. Again all you have to do is write them down."

Following this, all paper was removed from the table and a pad containing 12 sheets given to *S*. For those *Ss* who received no S-R pairs, this procedure followed completion of the first series. These instructions were read to all *Ss*: "I am going to show you some of these same syllables again. This time, on the top sheet of paper, write the first other nonsense syllable that you think of, fold the paper over immediately and do not look at it again. All you have to do is write down the first nonsense syllable that comes to mind or

that occurs to you, as long as it is not the same as the one on the card you are looking at; and as long as it is one of the syllables you have been writing down."

## Results and Discussion

A total of 810 associations was obtained from the group of 135 *S*s. Of these, 290 were S syllables, 403 were R syllables and 117 were extraneous syllables. Of the total number of R syllables, 175 were correct reproductions of the paired associate of the S syllable being presented. Although these frequencies cannot be subjected to a statistical test, the probability of obtaining this large a number of R, S, or correct syllables by chance, is, at most, very small. Furthermore, no *S* gave only extraneous syllables, and many gave none. Chi squares revealed that these extraneous syllables were not distributed in a manner significantly different from chance for any of the three main treatment variables and consequently these responses will not be considered further.

Each *S* was, therefore, given a score corresponding to each of the following: (*a*) number of R associations, (*b*) number of S associations, and (*c*) number of correct associations. Table 1 presents the means of the R and S scores for groups classified on the basis of R familiarity and frequency of S-R presentation.

With one exception, Table 1 shows a progressive increase in R scores and a decrease in S scores as a result of increasing R familiarity and increasing S-R occurrence. The exception to these trends occurs in the six S-R group least familiar with R syllables. This group produced more R associations and fewer S associations than expected from surrounding groups. The S scores are also observed to be higher than R scores in the zero S-R group, whereas R scores are higher than S scores in the groups exposed to S-R pairs. The exception to these trends occurs in the zero S-R group most familiar with R syllables where R scores are slightly higher

TABLE 1

Means and *SD*s of R and S Scores for R Familiarity and Frequency of Paired Presentation

| Scores | R Familiarity | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 2 | | 6 | | 10 | | Total | |
| | Mean | *SD* | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| R Scores S-R Trials: | | | | | | | | |
| 0 | 1.1 | 1.3 | 2.3 | 1.4 | 2.9 | 1.6 | 2.1 | 1.6 |
| 6 | 3.3 | 2.0 | 2.8 | 1.5 | 3.5 | 1.4 | 3.2 | 1.6 |
| 10 | 2.9 | 1.5 | 3.7 | 1.2 | 4.4 | 1.2 | 3.7 | 1.4 |
| Total | 2.4 | 1.9 | 2.9 | 1.5 | 3.6 | 1.5 | 3.0 | 1.7 |
| S Scores S-R Trials: | | | | | | | | |
| 0 | 4.5 | 1.5 | 2.7 | 1.8 | 2.4 | 1.6 | 3.2 | 1.9 |
| 6 | 1.3 | 1.3 | 2.0 | 1.2 | 1.7 | 1.7 | 1.7 | 1.4 |
| 10 | 1.5 | 1.6 | 1.9 | 1.5 | 1.1 | 1.1 | 1.5 | 1.3 |
| Total | 2.4 | 2.6 | 2.2 | 1.5 | 1.7 | 1.5 | 2.1 | 1.7 |

TABLE 2

ANALYSIS OF VARIANCE OF R AND S SCORES ON ASSOCIATION TEST TRIALS

| Source of Variation | *df* | *MS* | *F* |
|---|---|---|---|
| Between *Ss* | | | |
| Stimulus familiarity (A) | 2 | 1.233 | .74 |
| Response familiarity (B) | 2 | 2.211 | 1.33 |
| S-R frequency (C) | 2 | 1.300 | .78 |
| A × B | 4 | 1.477 | .90 |
| A × C | 4 | .611 | .37 |
| B × C | 4 | 1.516 | .91 |
| A × B × C | 8 | .336 | .20 |
| Error (Between *Ss*) | 108 | 1.657 | |
| Total Between *Ss* | 134 | | |
| Within *Ss* | | | |
| S and R scores (D) | 1 | 47.293 | 23.24*** |
| A × D | 2 | 19.471 | 9.57*** |
| B × D | 2 | 20.336 | 9.99*** |
| C × D | 2 | 66.382 | 32.62*** |
| A × B × D | 4 | 3.404 | 1.67 |
| A × C × D | 4 | 5.614 | 2.76* |
| B × C × D | 4 | 11.787 | 5.79*** |
| A × B × C × D | 8 | 6.251 | 3.07** |
| Error (within *Ss*) | 108 | 2.035 | |
| Total within *Ss* | 135 | | |
| Grand total | 269 | | |

* $P = .05$.
** $P = .005$.
*** $P = .001$.

than S scores. With the exception of the zero S-R group at the lower two levels of R familiarity, where *S* scores are higher than R scores, R scores are higher than S scores throughout Table 1.

Results of an analysis of variance performed on the R and S scores is summarized in Table 2.[3] The absence of significant variation arising from Between-*Ss* sources of variation suggests that *Ss* not only learned the syllables but produced learned syllables consistently at all levels of the three main treatment variables. Examination of Within-*Ss* variation, however, reveals significant differences between the R and S associations and all but one of the interactions between this variable and the three main treatment variables.

[3] The author wishes to express appreciation to W. L. Sawrey for his statistical design for this analysis and to Max Zwanziger for completing the computational procedures.

The over-all trend for *Ss* to give more R than S associations is reflected in the difference between R and S scores significant well beyond the .001 level. This suggests that *Ss* distinguished between the two classes of R and S syllables as a result of exposure to S-R pairs. Since two-thirds of the *Ss* were so exposed while one-third was not, a larger total number of R associations would be expected on this basis alone. This is reflected in the C × D interaction which is significant at well beyond the .001 level. In addition, the B × D interaction, also significant at the .001 level, reflects the trend toward higher R scores with increasing R familiarity. The B × C × D interaction, also significant at the .001 level, appears to result from the unusually high S scores observed at the lower level of R familiarity and zero S-R occurrence and the unusually high R scores observed at the highest level of R familiarity and S-R occurrence.

An increase in S scores with increasing S familiarity accompanied by high R scores at the lowest level of S familiarity is reflected in the A × D interaction, significant at the .001 level. The A × C × D interaction, significant at the .05 level, seems to reflect primarily the occurrence of unusually high S scores at the highest level of S familiarity in the absence of paired presentation and the unusually high R scores at the lowest level of S familiarity when pairing is present.

Since neither the A × B × C nor the A × B × D interaction approaches significance, the A × B × C × D interaction, significant at the .005 level, seems to reflect primarily the relative dominance of the effects of pairing over those associated with familiarity when these variables have opposite effects on the production of R and S scores. In spite of the facilitating effect on S scores associated with increasing S familiarity, S scores are higher than R scores only in those groups receiving no exposure to S-R pairs at the two lower levels of R familiarity. The facilitating effect of both R familiarity and S-R presentation on the production of R associations is sufficiently large to result in higher R scores everywhere else.

Since only S familiarity shows a facilitating effect on the production of S scores and all levels of S and R familiarity are present in the groups receiving no S-R pairs, the predominance of S scores even at the zero S-R level cannot be attributed to greater

familiarity with the S syllables alone. Apparently, some other factor, operating with little strength once exposure to pairs is introduced, exerts strong influence on associations in the absence of pairing. A reasonable accounting of this is provided by reference to a process which Skinner (1957) terms "formal prompting." Thus, although *S*s in the zero S-R group had no reason to divide the syllables into R and S classes, presentation of the S syllables as stimuli in the association test could function to prompt these syllables on subsequent trials. On the second test trial, for example, the probability of obtaining the first trial's S syllable as an association would be raised. On the third test trial, the probability of obtaining the second trial's S syllable would be raised. On the last test trial, all but one of the S syllables would have been so prompted. In *S*s receiving no S-R pairs, this could account for the proportionately greater strength of the facilitating effects of S familiarity when R familiarity is low and the balancing of effects of R familiarity when R familiarity is high.

Table 3 shows the distribution of correct associations for different levels of S-R occurrence. No *S* in the zero S-R group gave more than one correct association and more than two-thirds of *S*s in this group gave none. In groups exposed to syllable pairs, the number of correct associations increases as frequency of S-R presentation increases.

TABLE 3

DISTRIBUTION OF CORRECT ASSOCIATIONS FOR S-R FREQUENCIES

| S-R Frequency | No. Correct Associations | | |
|---|---|---|---|
| | 0 | 1 | 2 |
| 0 | 34 | 11 | 0 |
| 6 | 16 | 9 | 20 |
| 10 | 10 | 10 | 25 |

TABLE 4

MEDIAN TEST OF CORRECT ASSOCIATIONS FOR THE THREE MAIN TREATMENTS

| Source of Effect | *df* | $\chi^2$ |
|---|---|---|
| Stimulus familiarity (A) | 2 | 1.260 |
| Response familiarity (B) | 2 | .720 |
| S-R frequency (C) | 2 | 28.080* |
| A × B | 4 | 5.007 |
| A × C | 4 | 1.999 |
| B × C | 4 | .846 |
| A × B × C | 8 | 8.444 |
| Total | 26 | 46.356 |

* $P = .001$.

Since homogeneity of variance could not be established for correct scores, the nonparametric test of the medians described by Wilson (1956) for use in analyzing a three-factor design with replication was performed on these scores. A summary of this analysis is presented in Table 4. The trends noted in Table 3 are reflected in the chi square for the S-R frequency variable, which is significant well beyond the .001 level. Neither S familiarity, R familiarity nor any of the interaction effects approaches significance.

Although the trends associated with the S-R variable are in the expected direction, the absence of significant effects arising from variation in R or S familiarity is particularly interesting in view of the findings with respect to the two general classes of R and S syllables. The results of the analysis of correct scores is interpreted as providing further evidence that the effects of exposure to pairs of syllables takes precedence over both the effects of familiarity alone and any prompting effects inherent in the association method.

The importance of this study is seen as resting primarily in the presentation of a method for studying the acquisition of associations under experimentally con-

trolled conditions. The over-all results clearly indicate that variations in responses obtained by the association method reflect, systematically, variations in conditions of incidental, paired-associate learning of nonsense syllables. It does, therefore, seem reasonable to consider further the potential of this method for testing hypotheses derived from word association studies. Extension of the method to any study in which more than one R syllable is paired with a given S syllable, or any study in which more than one S syllable is paired with a given R syllable will, however, necessitate the confounding of either S or R familiarity with frequency of paired presentation. To provide a guide for designing and interpreting such extensions, the present study focused on exploring the systematic effect either R or S familiarity might be expected to produce independently of pairing. Findings that S familiarity is significantly related only to the frequency of S syllable associations and that both increased R familiarity and pairing were related to decreased frequency of S associations, as well as increased frequency of R associations, provides a basis for deciding how to introduce controls for familiarity in future use of the method. For example, if one focuses on studying variation in responses which can be unequivocally classified as R associations or "not S associations," it should be unnecessary to introduce independent control over S familiarity; but, provision for independent control over R familiarity should be included. If, on the other hand, more than one S syllable is paired with a given R syllable, to study the variation in S associations when the R syllable is used as the stimulus in the association test, independent control over both R and S familiarity should be introduced.

## Summary

This study presents a method for investigating, experimentally, the formation of associations through use of the association method as a test of incidental, paired-associate learning of nonsense syllables. Focus was placed on determining the relationships between variations in stimulus syllable familiarity, response syllable familiarity and frequency of pairing, and variations in associations. When associations could be viewed as a function of familiarity alone, syllables were given more frequently as associations as familiarity increased. Introduction of pairing tended to reinforce the tendency to give response syllables whether response syllables were more familiar or not and to decrease the tendency to give stimulus syllables whether these were more familiar or not. Pairing, further, was the only variable which related significantly to the production of correct associations.

## References

Bell, J. E. *Projective techniques.* New York: Longmans, Green, 1948.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

Jenkins, J. J., & Russell, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.*, 1952, **47**, 818–822.

McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.

Skinner, B. F. *Verbal behavior.* New York: Appleton-Century-Crofts, 1957.

Wilson, K. V. A distribution-free test of analysis of variance hypotheses. *Psychol. Bull.*, 1956, **53**, 96–101.

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received February 11, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 1, 1960

# PREDICTIONS OF SEQUENCES OF SUCCESSES AND FAILURES

CLINTON B. DE SOTO, EDMUND B. COLEMAN, AND PETER L. PUTNAM

*Johns Hopkins University*

In most research on levels of aspiration, *S* selects his level of aspiration from an achievement scale having many discriminable values (Lewin, Dembo, Festinger, & Sears, 1944; Simon, Shaw, & Gilchrist, 1954). The games he plays or the tests he takes have 10, 100, or more, possible scores. In many settings outside the laboratory, however, people are faced with an achievement scale having only two discriminable or acknowledged values, success and failure. A salesman either lands the customer or he doesn't. And in these cases, too, people set levels of aspiration in the sense that they expect either success of failure. As a salesman attempts to make a sale, he may expect either to succeed or to fail at it.

If people repeatedly face the same or similar tasks, succeeding a certain proportion of the time and failing the rest of the time, they will expect success a certain proportion of the time and failure the rest of the time. How is the proportion of trials on which they expect success related to the proportion of trials on which they actually succeed? The present article reports an attempt to answer this question experimentally.

It is proposed that a perfectly realistic *S* who succeeds on a constant proportion of trials will, in the long run, expect (predict) success on that proportion of trials. He will show *probability-matching*, in the terminology of the work on probability learning and guessing behavior. An attempt is made to support this proposal experimentally by showing that *S*s do match probabilities in predicting the successes of another, unknown person.

However, one of the consistent findings in studies of levels of aspiration has been that people usually set their level of aspiration somewhat higher than their previous level of achievement; they show positive goal discrepancies. By analogy, it is hypothesized that *S*s will in some degree overpredict their personal successes, i.e., will predict successes more often than they achieve them, rather than match probabilities.

## METHOD

The tasks given *S*s in this experiment are best described by the instructions given them.

*Predictions of another's successes.*—The instructions given *S*s in this control condition follow:

"This is a study of how well people can judge other people's successes and failures. You are to judge the results of a word association experiment that was recently taken by another student.

"One by one, he was given a series of words. He tried to give a correct association for each of these words according to rules which he was not told. He was told, after he gave each association, whether it was correct or incorrect. In other words, by noting whether his response was correct or not, he tried to learn the rules of association as he worked through the series of words.

"I will read to you the same series of words, in the same order, together with his associations. After each pair, I want you to tell me whether you think his association was correct or incorrect according to the rules (which I will not tell you, either). The rules remain the same throughout the experiment, and you may be able to discover some of them as we work through the series.

"Here are the steps we will go through:

1. I will read the word and his association, pair by pair.
2. After each pair, you are to say whether you think his association was correct or incorrect.
3. After your judgment, I will tell you whether his association was correct or incorrect."

The task was performed just as described in the instructions, *E* reading an actual transcript of the associations given by one of the experimental *S*s. There were no actual rules for associating words; *S* was told that the other *S* had succeeded or failed according to a programmed sequence which included 10 successes in each of five 30-trial blocks but was otherwise random.

*Predictions of personal successes.*—The instructions given *S*s in the experimental conditions follow:

"This is a word association experiment. On each of the cards in front of you is written one word. Certain other words can be correctly associated with each of these words, according to rules which I will not tell you. The rules will remain the same throughout the experiment, and you may be able to discover some of them as you work.

"You are to give one word in response to the word on each card. Your response should be a word you think may be correctly associated with the word on the card according to the rules. Before you give each response I want you to say whether you *expect* (not *hope*) it to be correct or incorrect. After you give your response I will tell you whether it actually was a correct or an incorrect one according to the rules.

"You should work for two goals: (*a*) to give correct responses whenever possible; (*b*) to judge as accurately as possible in each case whether your response actually will be correct or incorrect.

"These are the steps to go through for each card:

1. Pick up the card, turn it over, and read the word on the card *out loud.*
2. Decide on a word (*one* word) to give in response to the word on the card.
3. Say, 'I expect this to be correct,' or 'I expect this to be incorrect,' according to your best judgment.
4. Say your response word. (I will then say, 'That's correct,' if your response is a correct association according to the rules, or 'That's incorrect,' if it is not.)
5. Lay the card face down in a new pile. Pick up the next card and repeat the above steps."

The stimulus material consisted of 151 3 × 5-in. white cards on each of which was typed a single English word. The words used may be described by the following random sample: *because, comfort, gone, ink, nine, police, smooth, surface, true, why.*

The task was performed just as described in the instructions, except that there were no actual rules for associating words. The *E* told *S* he had succeeded or failed (had given a correct or incorrect response) according to a programmed sequence.

Each *S* controlled the pacing of his trials. The mean time taken to complete 151 trials was 32.5 min.; the range was 15 to 60 min. Time taken to complete the experiment was not related to the experimental conditions which will be discussed or to individual differences on other scores.

Five sequences of successes and failures were programmed. The essential difference between them was the proportion of successes, which was $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, or $\frac{5}{6}$ in the various cases. These proportions held true exactly within 30-trial blocks of each sequence; that is, in successive 30-trial blocks *S* had 5, 10, 15, 20, or 25 successes, the same number in each block according to the sequence *E* was giving him. Trial 151 was a success in all sequences. Subject to these restrictions, the sequences of successes and failures were random.

The experimental conditions varied only in which of the five sequences of successes and failures was administered. In the rest of this article, they will be referred to as the $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{2}{3}$, and $\frac{5}{6}$ conditions.

*Subjects.*—The *S*s were 96 male undergraduates from an introductory psychology course at Johns Hopkins, 16 in each condition.

## Results

*Predictions of another's success.*—Table 1 shows by 30-trial blocks the mean number of trials on which *S*s in the control condition predicted successes, together with the standard deviation of their scores for each block. These *S*s were making predictions for another, unknown person who actually succeeded on 10 of the 30 trials in each block. For a while, *S*s predicted somewhat more suc-

TABLE 1

PREDICTIONS OF ANOTHER'S SUCCESSES

| Trials | Mean | *SD* |
|---|---|---|
| 1–30 | 13.6 | 3.5 |
| 31–60 | 13.9 | 3.9 |
| 61–90 | 11.9 | 4.8 |
| 91–120 | 10.7 | 3.8 |
| 121–150 | 9.6 | 3.1 |

cesses for him than he attained, but in the latter blocks they predicted for him about the number he actually attained, showing probability-matching.

*Predictions of personal successes.*—Table 2 shows corresponding data for each of the experimental conditions. These *S*s were making predictions of their own successes while actually succeeding on 5, 10, 15, 20, or 25 trials in each 30-trial block.

The *S*s did not optimistically predict success all the time at any stage in any condition, even in the condition in which they were succeeding on $\frac{5}{6}$ of the trials. They predicted success on a proportion of trials which was positively related to the proportion of trials on which they actually succeeded—the more successes, the more predictions of successes. This result is analogous to the usual finding that level of aspiration is an increasing function of level of previous achievement. And it is like the results of probability-learning experiments in general in that frequency of prediction of an event is always an increasing function of the frequency of its occurrence.

In most probability-learning experiments, however, *S*s approximate probability-matching over many trials, just as they did in predicting another's successes in the present experiment. The *S*s did not show probability-matching in predicting their own successes. If probability-matching is taken as a baseline, *S*s persistently overpredicted success in most of these conditions; they predicted it far more often than they obtained it. This overprediction was greatest for the least successful conditions, lessening as the proportion of actual successes increased, until in the $\frac{5}{6}$ condition *S*s actually slightly underpredicted success.

Predictions of success in the $\frac{1}{6}$ condition show a decided drop over trials, although even in the last blocks *S*s on the average predict almost twice as many successes as they obtained.

Individual differences as reflected in the standard deviations in Table 2 are somewhat larger than in typical probability-learning experiments (Estes & Straughan, 1954), but it was rare for *S*s to approach the extremes of predicting all successes or all failures. Only one *S* in the entire experiment gave the same prediction on all 151 trials, an *S* in the $\frac{1}{6}$ condi-

TABLE 2

PREDICTIONS OF PERSONAL SUCCESSES

| Trials | Cond. $\frac{1}{6}$ | | Cond. $\frac{2}{6}$ | | Cond. $\frac{3}{6}$ | | Cond. $\frac{4}{6}$ | | Cond. $\frac{5}{6}$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| 1–30 | 14.9 | 7.0 | 17.6 | 4.2 | 17.6 | 4.7 | 23.1 | 5.4 | 23.8 | 4.4 |
| 31–60 | 13.4 | 6.7 | 17.8 | 3.2 | 19.2 | 5.8 | 21.7 | 6.5 | 23.9 | 4.7 |
| 61–90 | 10.4 | 7.5 | 16.6 | 5.7 | 20.4 | 5.1 | 22.6 | 5.7 | 23.4 | 6.0 |
| 91–120 | 9.8 | 7.2 | 16.2 | 5.3 | 17.9 | 7.1 | 22.1 | 6.1 | 23.9 | 5.1 |
| 121–150 | 9.1 | 6.9 | 19.4 | 4.1 | 19.2 | 5.7 | 22.3 | 6.4 | 25.5 | 4.9 |

tion who predicted failures throughout. During the last 30-trial block, four *S*s in the $\frac{5}{6}$ condition and one *S* in the $\frac{1}{2}$ condition predicted success on every trial.

*Mean transition probabilities.*—Some empirical transition probabilities were obtained for the $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{2}{3}$ conditions. For each of the 48 *S*s in these conditions, the following proportions were calculated.

$P(C_S|C_S,C_E)$—the proportion of trials on which *S* predicts he will be correct after having on the preceding trial predicted he would be correct and then been told by *E* that he was indeed correct.

$P(C_S|C_S,I_E)$—the proportion of trials on which *S* predicts he will be correct after having on the preceding trial predicted he would be correct but then been told by *E* that he was incorrect.

$P(C_S|I_S,C_E)$—the proportion of trials on which *S* predicts he will be correct after having on the preceding trial predicted he would be incorrect but then been told by *E* that he was correct.

$P(C_S|I_S,I_E)$—the proportion of trials on which *S* predicts he will be correct after having on the preceding trial predicted he would be incorrect and then been told by *E* that he was indeed incorrect.

These proportions were averaged over the 48 *S*s to yield four mean transition probabilities, which are as follows:

$$P(C_S|C_S,C_E) = .69.$$
$$P(C_S|C_S,I_E) = .62.$$
$$P(C_S|I_S,C_E) = .63.$$
$$P(C_S|I_S,I_E) = .65.$$

A Friedman analysis of variance by ranks was performed on the four sets of proportions, yielding a $\chi^2_r$ of 12.73, significant at the .01 level with 3 *df*. This result justifies a conclusion that at least the larger differences among the figures above are reliable.

A useful aid to interpreting these figures is to compare the averages of pairs of them. There are three ways this can be done. One is to compare the first two with the last two. The average of $P(C_S|C_S,C_E)$ and $P(C_S|C_S,I_E)$ is .655. The average of $P(C_S|I_S,C_E)$ and $P(C_S|I_S,I_E)$ is .64. This comparison indicates *S*s, on the average, are slightly perseverative in their predictions of correct and incorrect. They tend to predict success on a given trial if they predicted success on the preceding trial; they tend to predict failure on a given trial if they predicted failure on the preceding trial.

Another comparison is the first and third figures vs. the second and fourth. The average of $P(C_S|C_S,C_E)$ and $P(C_S|I_S,C_E)$ is .66. The average of $P(C_S|C_S,I_E)$ and $P(C_S|I_S,I_E)$ is .64. This comparison indicates *S*s, on the average, tend to predict on a given trial the level of achievement, success or failure, that they attained on the preceding trial.

The final comparison is the first and fourth figures vs. the second and third. The average of $P(C_S|C_S,C_E)$ and $P(C_S|I_S,I_E)$ is .67. The average of $P(C_S|C_S,I_E)$ and $P(C_S|I_S,C_E)$ is .625. This comparison, yielding the strongest contrast of the three, indicates that *S*s tend to predict success on a given trial if their prediction of success or failure on the preceding trial was *accurate*.

*Ss' reports on their hypotheses.*—After the experiment, *S*s were asked if they had discovered the rules by which words could correctly be associated with the words on the cards. Of the 80 *S*s only one failed to report

a hypothesis about the rules. Two *Ss* directed their hypotheses to the actual sequence of successes and failures, while all the rest expressed hypotheses about the rules of word associations, hypotheses which were necessarily vague and qualified.

## Discussion

The fact that *Ss* showed probability-matching in predicting the successes of another, unknown person strongly supports the proposal that a perfectly realistic *S* would show probability-matching in predicting his own successes. Also, the data generally support the hypothesis that *Ss*, in actuality, tend to overpredict their personal successes. However, the degree of overprediction is surprisingly great in the least successful conditions and surprisingly slight in the most successful conditions.

It seems worthwhile to see how well the lessening of overprediction with increasing success can be accounted for by the traditional explanatory concepts of level-of-aspiration research. It is generally accepted that what makes people depart from the best realism of which they are capable in setting levels of aspiration is the varying attractiveness (or valence or utility) of varying levels of achievement. The greater attractiveness of higher, more difficult levels makes a person aspire to a level somewhat above that which past experience and other information objectively would dictate, the size of the so-called positive goal discrepancy depending on situational requirements for realism (Festinger, 1942). In the case of a two-level achievement scale, it would seem that the perceived difficulty, hence attractiveness, of achieving successes must depend primarily on the relative frequency with which *Ss* achieve them. The fewer the successes *S* achieves, the more worthwhile and attractive they are, and the stronger is his tendency to depart from realistic probability matching to overpredict them. This formulation ignores the possibility that as the task becomes more difficult and success becomes more attractive, failure may also become more acceptable, less unattractive, but it is probable that the attractiveness of failure is less labile than that of success and is therefore a minor consideration.

The mean transition probabilities reported indicate that average recency effects were rather slight, which was to be expected from other probability-learning studies in which *Ss*' bases for predictions were directed away from the sequence of events itself to other factors determining the events (Goodnow, 1955). The evidence for at least three recency effects, however slight, does indicate that detailed specification of *Ss*' predictions would require consideration of many determinants. It is likely too that there are individual differences in such recency effects. For example, there were indications in the data that degree of overprediction of successes was associated with degree of response perseveration.

It should be emphasized that the observed large departures from the supposed realism of probability matching were obtained despite admonitions to *Ss* in the instructions to be realistic. It is known that *Ss* given such admonitions when setting levels of aspiration on a many-valued achievement scale show only slight goal discrepancies (Festinger, 1942). Apparently such realism is not so easily obtained on a two-category achievement scale. Now the work on probability-learning has established that people can discriminate probabilities in the sense that they predict events with about the same frequency as they occur. They have the discriminative capacity to be realistic in predicting successes on a two-level scale, just as on a many-level scale, but are not so easily compelled to be realistic, perhaps because their lack of realism is less conspicuous to themselves and to others. It appears that in promoting realism in people's aspirations, endeavors, and self-evaluations, it is very important to provide them with feedback information more detailed than mere word that they have succeeded or failed.

## Summary

This experiment was concerned with *Ss*' predictions of successes and failures on a word-association task where the actual sequence of successes and failures had been programmed by *E*.

In predicting the successes of another, unknown, person who succeeded on $\frac{1}{3}$ of the trials, *Ss* eventually showed probability matching, predicting successes at about the frequency with which they occurred.

In predicting their personal successes on the task, the proportion of trials on which *Ss* predicted success was an increasing function of the proportion of trials on which they actually succeeded, but they persistently predicted more successes than they actually obtained. This overprediction of successes was considered to be analogous to positive goal discrepancies in the setting of levels of aspiration on a many-valued achievement scale. Since *Ss* were admonished to be realistic in their predictions, it was concluded that realism in predictions of personal successes and failures (identified as probability matching) is not as easily obtained as realism in the setting of levels of aspiration on a more finely differentiated achievement scale. The higher the proportion of trials on which *Ss* were successful, however, the less they overpredicted, and *Ss* who succeeded on $\frac{1}{3}$ of the trials did not overpredict successes.

Examination of first-order transition probabilities in their predictions showed that *Ss* were more likely to predict success if they had succeeded on the previous trial than if they had not, they were more likely to predict success if they had predicted success on the previous trial than if they had not, and they were more likely to predict success if they had predicted accurately on the preceding trial than if they had not.


## References


Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Festinger, L. Wish, expectation, and group standards as factors influencing level of aspiration. *J. abnorm. soc. Psychol.*, 1942, **37**, 184–200.

Goodnow, J. J. Response-sequences in a pair of two-choice probability situations. *Amer. J. Psychol.*, 1955, **68**, 624–630.

Lewin, K., Dembo, T., Festinger, L., & Sears, P. S. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. I. New York: Ronald, 1944. Pp. 333–378.

Simon, J. R., Shaw, M. E., & Gilchrist, J. C. Some effects of prearranged performance scores upon the level of aspiration. *J. exp. Psychol.*, 1954, **47**, 10–12.


(Received February 12, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 1, 1960

# GRADIENTS OF GENERALIZATION IN SECONDARY REINFORCEMENT

BRUCE O. BERGUM [1]

*Northwestern University*

Studies concerned with the generalization of secondary reinforcement (Denny, 1948; Ehrenfreund: 1949, 1954; Saltzman, 1950) typically involve comparison of two groups, one with similar, the other with less similar, secondary reinforcing cues. These studies have established the fact that stimulus generalization is a variable in secondary reinforcement, but they do not provide evidence as to whether there is a gradient in such generalization. The purpose of the present study is to determine whether such a gradient exists and, if one does, to determine its general form.

The present study is modeled somewhat after those of Saltzman (1949; 1950). Basically, it involves training in a runway to establish the secondary reinforcing value of a given stimulus, followed by testing for generalization in a new learning situation (U maze) with five physically defined levels of similarity to the original training stimulus.

## METHOD

*Subjects.*—The *S*s were 100 male albino rats of the Sprague-Dawley strain, ranging in age from 60 to 75 days.

*Apparatus.*—The apparatus consisted of a runway, a single-choice U maze, an interchangeable starting box, one black and one white goal box for use with the runway, and six goal boxes for the maze. All of the apparatus (except the starting box which was 9-in. square) was 6 in. wide and 8 in. deep in interior dimensions.

The starting box, runway, and U maze were painted with alternating ½-in. wide black and white vertical stripes to make them distinctive from the goal boxes, and to minimize generalization from the goal boxes to the various alleys.

The six maze goal boxes varied in shade from white through black in approximately subjectively equal intensity intervals as determined by three human judges. Illuminometer readings in foot-candles for the six shades *in situ* were: 29.2 for the white, 14.5 for the light gray, 6.8 for the medium light gray, 3.6 for the medium dark gray, 1.9 for the dark gray, and 1.0 for the black.

*Procedure.*—For six days prior to training all *S*s were placed on a feeding schedule consisting of 10 gm. of dry pellet food in the home cage for 1 hr. in every 24. During this preliminary period *S*s were tamed and familiarized with both the dry pellets and the various parts of the apparatus. The reward pellets used in the runway were standard .05 gm. laboratory pellets.

On Days 1 and 2 of the experiment proper, each *S* received 40 trials in the runway on each day. On the first trial each day *S* ran to a black-positive (white for half the animals) goal box where he received a single pellet of food. Following this trial, *S* received each day an additional 24 reinforced trials running to the positive goal box, randomly alternated with seven nonreinforced trials to the positive goal box and eight nonreinforced trials to the negative goal box. At the completion of a run *S* was allowed to remain in the goal box for 10 sec., after which he was placed in the home cage for 15 sec. preceding the next trial.

On Day 3 of training *S*s were randomly divided into five groups for training in the U maze. For the 20 *S*s in Group I the "correct" (i.e., secondarily rewarded) maze goal box was the same color as the box in which *S* was reinforced during runway training. For Group II the correct maze goal box was the one most similar in color to the rein-

[1] Based upon a dissertation submitted to the Graduate School, Northwestern University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to express gratitude to his adviser, Carl P. Duncan, and to Donald J. Lewis and John W. Cotton for their assistance in carrying out the study.

forced runway box. For Group III the correct maze box was that second most similar, for Group IV the correct box was that third most similar, for Group V the correct box was that least similar to the positive runway goal box. Therefore, for Group V the correct maze goal box was most similar, but not identical, to the negative runway box. In addition, the various counterbalancing procedures were such as to permit the testing of gradients both from white to black and from black to white.

For half of the *S*s the correct goal box appeared on the left side of the maze, and for the remaining half, on the right. For all *S*s the "incorrect" maze goal box was the never reinforced (black or white) goal box used in the runway on the previous two days.

All *S*s were given one forced trial to the correct goal box just preceding the first test trial in the U maze, and all *S*s then received 15 trials in the U maze with a modified correction technique; i.e., *S*s were allowed to correct the turn they made at the choice point as long as they did not enter a goal box. However, once they entered either goal box the door was dropped behind them and they were kept in the box for 10 sec. The *S*s received *no* primary reward in either goal box at any time during the maze training. Following each trial *S* was placed in the home cage for 20 sec.

The response measures during maze training were (*a*) reciprocal running times measured from the time the door was raised on the starting box until *S* made the *first* turn at the choice point on each trial, (*b*) frequency with which the first turn was correct, and (*c*) frequency with which the goal box entry was correct.

## Results

Analyses of the maze data will be presented for each of the three response measures separately.

*Number of correct turns.*—The score for each *S* for this measure was the total number of correct turns made during all 15 trials in the maze. The mean number of correct turns for each of the 10 groups ($N = 10$ per group) is shown in Table 1. In Table 1 and in subsequent similar tables "reinforced color" indicates the color of the positive goal box during runway training, "similarity level" indicates degree of similarity of the correct maze goal box to the reinforced color.

TABLE 1

Mean Number of Correct Turns over all 15 Trials in the U maze

| Reinforced Color | Similarity Level | | | | | Mean |
|---|---|---|---|---|---|---|
| | I | II | III | IV | V | |
| White | 9.90 | 9.40 | 7.90 | 8.10 | 7.90 | 8.64 |
| Black | 10.50 | 8.80 | 9.10 | 7.80 | 6.60 | 8.56 |
| Mean | 10.20 | 9.10 | 8.50 | 7.95 | 7.25 | |

The grand means in the bottom row of Table 1 show a systematic trend without reversal in the effects of the similarity variable with the most similar U-maze goal box yielding the greatest number of correct turns and the least similar goal box yielding the smallest number. The color of the originally reinforced runway goal box does not appear to have differentially affected performance.

The data were analyzed by a $2 \times 5$ factorial analysis of variance in which the two dimensions were (*a*) reinforced color, and (*b*) similarity level. The results of the analysis, presented in Table 2, show that the similarity variable was highly significant. Bartlett's test for homogeneity of variance yielded a $\chi^2$ of 12.99 for 9 *df*, which is not significant.

The systematic effect of the similarity variable is shown by the lower curve in Fig. 1, in which the mean

TABLE 2

Analysis of Variance of Number of Correct Turns

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Reinforced color | 1 | 0.160 | |
| Similarity | 4 | 25.325 | 7.43* |
| C × S | 4 | 4.885 | 1.43 |
| Within | 90 | 3.411 | |

* $P < .001$.

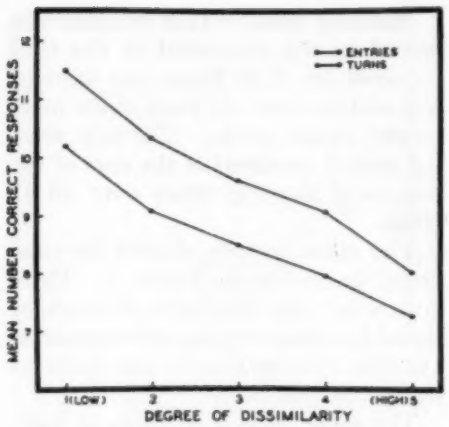


FIG. 1. Mean correct turns and mean correct goal box entries in the maze as a function of degree of dissimilarity in brightness of the reinforced runway goal box to the correct maze goal box.

number of correct turns is plotted against similarity level. With the exception of a slight depression at Level V below the chance level of performance (7.5 mean correct turns), the empirical curve is very regular. In Fig. 1 the similarity levels are equally spaced on the baseline for human judges. If the baseline were plotted in log foot-candles of goal box luminance, the curves would look much the same.

To examine the possible improvement in performance through the course of experience in the maze, the total numbers of correct turns during the first five, second five, and last five trials were determined for each *S* at each similarity level. All similarity groups appear to show improvement in performance with practice most of the change occurring in the first 10 trials. The differential effects of the similarity variable appear to establish and maintain themselves almost from the beginning of testing.

In order to determine whether the apparent change in performance shown in the center curve of Fig. 2 was significant, an analysis of variance for similarity level and stage of practice, summarized in Table 3, was performed on the 15 means. Table 3 shows that, as before, the similarity effects were significant beyond the 1% level. Of more immediate interest, the stages-of-practice variable was also significant beyond the 1% level indicating a significant improvement in performance in the course of U-maze training.

TABLE 3

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT TURNS AT VARIOUS STAGES OF PRACTICE

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Similarity | 4 | 0.835 | 41.75* |
| Stages | 2 | 1.190 | 59.50* |
| Sim. × Stage | 8 | 0.020 | |

* $P < .01$.

*Number of correct goal box entries.*—On this measure each *S*'s score consisted of the total number of correct goal box entries made during all 15 trials. The mean number of entries is presented in Table 4. These means, also shown in the upper curve of Fig. 1, parallel almost exactly those for the turns data except that they tend to average slightly more than one response higher.

The entries data were analyzed by the same type of analysis of variance

TABLE 4

MEAN NUMBER OF CORRECT ENTRIES OVER ALL 15 TRIALS IN THE U MAZE

| Reinforced Color | Similarity Level | | | | | Mean |
|---|---|---|---|---|---|---|
| | I | II | III | IV | V | |
| White | 11.60 | 10.30 | 9.70 | 9.20 | 8.60 | 9.88 |
| Black | 11.40 | 10.40 | 9.50 | 8.90 | 7.40 | 9.52 |
| Mean | 11.50 | 10.35 | 9.60 | 9.05 | 8.00 | |

used on the turns data (Table 2). No table is presented for this analysis since the results were almost identical to those for turns. Neither original reinforced color nor the interaction was significant. Similarity was significant beyond the .001 level ($F = 8.16$ for 4 and 90 *df*). The Bartlett test for homogeneity of variance was not significant ($\chi^2 = 4.49$ for 9 *df*).

The top curve of Fig. 2 shows the means for the entries data broken down into thirds of practice. The results are much the same as for the turns data except that in this case the means tend to fall at or above chance, probably because an alternation tendency should have less effect on entries than on turns. With the correction procedure used, *S* could still enter the correct goal box even after making an incorrect turn.

The same type of analysis of variance used for the turns data was done on the means for entries. The results were much the same as those shown in summary Table 3; similarity was significant at the 1% level ($F = 17.06$ for 4 and 8 *df*), stages of practice was significant at the 1% level ($F = 10.15$). Thus, a significant improvement in performance also took place in terms of the entries measure.

Except in reflecting a slightly higher level of performance, the results for the number of correct entries parallel exactly those for the turns data. The tendency to yield a higher level of performance may indicate that the entries measure is somewhat less affected by irrelevant turn biases than is the turning response, and probably is a more sensitive measure of secondary reinforcing effects. Both measures, however, apparently demonstrate equally well the general form of the generalization gradient.

*Running times.*—This measure was scored as the reciprocal of the time required by *S* to leave the starting box and traverse the stem of the maze to the choice point. The raw score for each *S* consisted of the sum of the reciprocal running times over all 15 trials.

The mean reciprocal time for each group is shown in Table 5. These data reflect the same over-all gradient found for the preceding two measures, but the relationship is not quite as smooth in this case.

The same type of analysis of variance (Table 2) used on the other two measures was applied to the reciprocal running times. The results were essentially the same as those for turns and entries. Neither reinforced goal box color ($F < 1.00$) nor the interaction ($F = 1.51$) was significant. Similarity, however, was significant beyond the .001 level ($F = 7.20$). Bartlett's homogeneity of variance test yielded a nonsignificant $\chi^2$ of 15.18 for 9 *df*.

As in the case of turns and entries, the running time data were broken down into three stages of training, as shown again in Fig. 2. Of particular interest is the fact that, in contrast to the progressive improvement shown by the turns and entries data, running times display a typical extinction curve form.

TABLE 5

MEAN RECIPROCAL RUNNING TIMES OVER ALL 15 TRIALS IN THE U MAZE

| Reinforced Color | Similarity Level | | | | | Mean |
|---|---|---|---|---|---|---|
| | I | II | III | IV | V | |
| White | 10.35 | 11.80 | 9.28 | 8.35 | 7.28 | 9.41 |
| Black | 13.17 | 12.37 | 5.99 | 8.39 | 6.19 | 9.22 |
| Mean | 11.76 | 12.08 | 7.63 | 8.37 | 6.73 | |

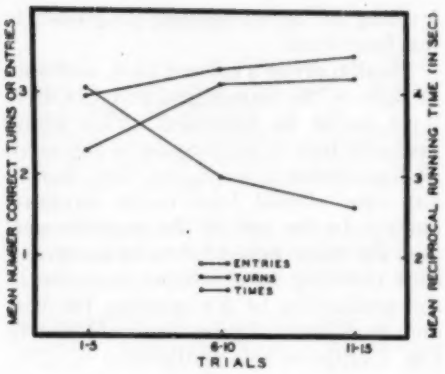


FIG. 2. Mean correct turns, mean correct goal box entries, and mean reciprocal running times in terms of three stages of practice in the maze.

The same type of analysis of variance used to analyze stages of practice in turns and entries data was performed on the 15 means for similarity and stage of practice. As expected, the similarity effects were significant beyond the 1% level ($F = 7.16$ for 4 and 8 *df*). The *F* for stages of training was 11.06, which for 2 and 8 *df* is significant beyond the 1% level. Thus, it appears that significant extinction of the running response occurred in the course of U-maze training. This presents the interesting paradox of *S*s running slower and slower while at the same time showing significant improvement in the performance of correct turns and entries. This relationship is shown graphically in Fig. 2.

Except for showing a decrease in performance level, the evidence from the running times data follows the same general trend as that found with the turns and entries data. The conclusion that secondary reinforcing stimuli generalize systematically as a function of their objective similarity to the originally reinforced stimulus is strongly supported.

## DISCUSSION

The purpose of this experiment was to study the generalization of secondary reinforcement as a function of stimulus similarity along an intensity continuum, and to determine the general form of the gradients associated with three response measures.

The results of the study were unambiguous. Response strength was directly related to the degree of similarity of the secondary reinforcing stimulus to the originally reinforced stimulus. These results are in conformity with what would be expected on the basis of a large body of evidence relating to stimulus generalization. In terms of gradient forms, both turns and entries yielded approximately straight-line gradients, but while end-point differences for the running time measures were significant, the form of the gradient was not as clear.

The significant improvement in performance for the turns and entries data is open to alternative interpretations as a result of the specific U-maze procedure employed in this study. Since the procedure involved a preliminary forced trial to the correct side, it is possible that this may have led to an increased probability of an incorrect turn on the first test trial. The slight depression below the chance level of 2.5 correct turns in the first five trials, apparent in Fig. 2, lends some support to this notion. The improved performance over trials could thus simply represent recovery from this alternation tendency. Dember and Fowler (1958) present some of the arguments for the alternation interpretation of performance increments and a summary of the relevant literature.

The alternative interpretation would be that the improved performance over trials found in the present study represents learning. If the improvement is so interpreted, the results for all three response measures might be explained in terms of the more recent theoretical writings of Spence (1956). In his formulation, *N*, number of S-R occurrences, becomes the single effective variable

determining associative strength, while an indeterminate number of other variables, including $N$, affect the intervening variable $K$, a motivational factor. Secdary reinforcement is the association of $r_g$'s with stimuli proximate to the primary reinforcer by a process of classical conditioning. Elicitation of these fractional responses results in heightened tension which contributes to the general drive level. Thus, secondary reinforcement is a motivational variable consisting of classically conditioned $r_g$'s which follow the traditional laws of conditioning.

This formulation would lead to certain predictions in the context of the present study. Thus, if $r_g$ is classically conditioned, we would expect that in the course of U-maze training, where primary reinforcement never occurred, $r_g$ would tend to extinguish, motivation would decrease, and $S$ would tend to run more and more slowly in the course of training. As indicated in Fig. 2, this prediction was borne out.

Presumably, $r_g$ is elicited to a greater or lesser extent in the correct goal box of the U maze, and as practice continues, stimuli at progressively earlier points come to elicit $r_g$. Elicitation of these $r_g$'s would increase $K$ and thus increase the motivational level of $S$. Given the secondary reinforcing effects of having entered the "correct" goal box at least once, the difference in $H$ between "correct" and "incorrect" habits would be multiplied by a factor greater than unity, increasing the probability of correct turns. Differences in the probability of eliciting $r_g$ in the goal box as a function of similarity to the originally reinforced runway goal box would account both for the over-all differences in running times among groups and the differences in turns. With lower probabilities of eliciting $r_g$, less $K$ would be generated and would result in a decrease in the probability of a correct turn. The effect of this mechanism would, of course, be greatest early in learning, before it had extinguished, with the effect leveling off as $r_g$ became progressively less functional.

Finally, given a correct turn, and once in sight of the correct goal box, $S$'s drive level would be increased. This would probably lead to an increase in the speed of locomotion toward the box, but in any case should lead to its eventual entry. In the case of the negative goal box, the drive would fail to be increased, thus resulting in a relative decrease in the probability of $S$'s entering the correct goal box. Again, the evidence in Fig. 2 supports this prediction.

## Summary

The experiment was designed to study the generalization of secondary reinforcement as a function of stimulus similarity along an intensity continuum, and to determine the general form of the gradients associated with three response measures.

Five groups of white rats (20 per group) received 80 trials in a runway on two consecutive days, running to a partially reinforced goal box on 64 trials and to a nonreinforced goal box on 16 interspersed trials. The goal boxes were black or white. On the third day of training all $S$s were given 15 trials in a U maze in which the nonreinforced goal box appeared at the end of one arm and a goal box bearing one of five degrees of similarity in brightness to the originally reinforced goal box, appeared at the end of the other arm. The points along the similarity continuum differed by equal appearing intervals according to human judges. The $S$s were never given food reinforcement in either goal box in the U maze.

Approximately straight-line generalization gradients were shown by two of the three response measures, correct turns, and correct goal box entries, and these gradients yielded statistically significant end-point differences in all cases. In the case of reciprocal running times the end-point differences were significant, but the form of gradient was not as clear. There was no difference between generalizing from white to black and black to white. The data were further analyzed in terms of stage-of-practice in the U maze and it was found that turns and entries showed significant improvements in performance, while simultaneously the running time showed a significant decrease in performance level.

## References

DEMBER, W. N., & FOWLER, H. Spontaneous alternation behavior. *Psychol. Bull.*, 1958, **55**, 412–428.

DENNY, M. R. The effect of using differential end boxes in a simple T-maze learning situation. *J. exp. Psychol.*, 1948, **38**, 245–249.

EHRENFREUND, D. Effect of a secondary reinforcing agent in black-white discrimination. *J. comp. physiol. Psychol.*, 1949, **42**, 1–5.

EHRENFREUND, D. Generalization of secondary reinforcement in discrimination learning. *J. comp. physiol. Psychol.*, 1954, **47**, 311–314.

SALTZMAN, I. J. Maze learning in the absence of primary reinforcement: A study of secondary reinforcement. *J. comp. physiol. Psychol.*, 1949, **42**, 161–173.

SALTZMAN, I. J. Generalization of secondary reinforcement. *J. exp. Psychol.*, 1950, **40**, 189–193.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

(Received February 15, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 1, 1960

# RETENTION OF RESPONSES TO STIMULUS CLASSES AND TO SPECIFIC STIMULI [1]

KENNETH E. LLOYD [2]

*Ohio State University*

It is frequently assumed that retention of meaningful material is superior to that of nonmeaningful material. While precise definitions of meaningfulness are elusive, it is often stated that prose, concepts, or substance material meet the criteria of meaningfulness. Comparisons of retention of such materials with that of nonsense syllables have been made, but rather than specifying the measurable variables that contribute to the observed retention differences, explanations have centered around the qualitative aspects of the learning materials. One variable that could influence such results is the number of response units that must be made by *Ss* who learn different kinds of material.

For example, Reed (1946) concluded that retention of concepts was far superior to retention of nonsense syllables. The *Ss* in his concept group associated six nonsense syllable responses to 42 verbal stimuli which could be grouped into six concepts with seven instances of each. The retention of these *Ss* was compared with that of Ebbinghaus (1913) who usually learned eight lists of 13 syllables in one session; i.e., the retention of six responses was compared with that of 104 responses. Katona (1940) requested *Ss* to learn a series of 24 numbers, but varied his instructions to different groups in such a way that discovery of a rule, by some of the groups, reduced the number of separate items that had to be remembered from 24 to 4. The superior retention of these concept groups may have been a function of the relatively fewer number of responses learned by them, rather than a function of the meaningfulness or organization of the material. The present experiment controlled this possibility by equating the number of response units which *Ss* learned.

The hypothesis tested was that there would be no difference in retention between *Ss* who acquired responses to specific stimuli and *Ss* who acquired responses to classes of stimuli if the number of response units learned was the same in both situations. Responding to specific stimuli means that a given response is emitted only in the presence of a given stimulus; responding to classes of stimuli means that a given response is emitted in the presence of any instance of a class of stimuli.

No presumption is made that the latter condition is necessarily more meaningful than the former. Nevertheless, responding to classes of stimuli should involve a mediating response while responding to specific stimuli need not. It is possible to determine empirically whether or not *Ss* in the class groups are employing a mediating word. If *Ss* who learn arbitrary responses to instances of a stimulus class are presented with different in-

[1] This article is based on a dissertation submitted to the Graduate School of the Ohio State University in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express his appreciation to Paul M. Fitts whose stimulation greatly facilitated this work.

[2] The author is now at Washington State University.

stances of this class, their performance should exceed that of *Ss* who have learned responses only to specific words. When the new instance is presented, *S* makes the mediating response which then sets the occasion for the arbitrary response. If the response has been made only to a specific stimulus, then no mediating response need be employed, and *S* is not set for a new instance.

For example, one group could learn $R_1$ to the words *priest*, *monk*, and *deacon*, while another group learned the same response to the one word *priest*. If both groups are later presented with the word *bishop*, then the former group should emit $R_1$ more quickly than the latter. Positive transfer effects would undoubtedly operate in both conditions, but the class groups should benefit more if they have employed a mediating word such as *churchmen*.

A transfer task was used in this study as a means of testing the proposition that one group responded to classes of stimuli while the other group responded to specific stimuli. No retention differences were expected, however, since the number of responses in the two conditions was identical.

## Method

*Design.*—The design consisted of three steps, an original learning (L-I), a recall trial, and a transfer task (L-II). During L-I two specific groups (experimental and control), and two class groups (experimental and control) learned eight number responses either to eight specific words, or to 24 words which could be categorized into eight classes, i.e., three instances of each class. After the recall trial, the experimental specific group (Group ES) learned a second list containing eight different specific words from the same classes used in L-I. The control specific group (Group CS) transferred to eight specific words from entirely different classes. The experimental class group (Group EC) and its control (Group CC) transferred to lists containing three new instances of the old classes, and of new classes, respectively. The same eight responses were used in both L-I and L-II. Table 1 illustrates the design with words taken from two classes of stimuli.

TABLE 1

STIMULUS WORDS AND THEIR RESPONSE NUMBERS FOR THE FOUR GROUPS DURING THE ORIGINAL LEARNING AND THE TRANSFER TASKS (THE WORDS ARE INSTANCES OF TWO CLASSES)

| Group | Learn I | | Learn II | |
|---|---|---|---|---|
| | Stim. | Resp. | Stim. | Resp. |
| ES | Priest[a] | 1 | Bishop | 1 |
| CS | Bishop[a] | 1 | Cottage | 1 |
| EC | Priest | 1 | Bishop | 1 |
| | Monk | 1 | Missionary | 1 |
| | Deacon | 1 | Chaplain | 1 |
| CC | Bishop | 1 | Cottage | 1 |
| | Missionary | 1 | Garage | 1 |
| | Chaplain | 1 | Skyscraper | 1 |

[a] *Ss* in Groups ES and CS were divided into six subgroups so that all the instances of a class could serve as specific words; e.g., during L-I seven *Ss* in Group ES saw only the word *priest*, seven other *Ss* saw only the word *monk*, etc. This counterbalancing produced no performance differences.

*Procedure.*—The 72 stimulus words were selected on the basis of their class membership and their frequency of use (Thorndike & Lorge, 1952). Within each class there were both relatively frequent and relatively infrequent words. Eight preliminary *Ss* were asked to sort these words plus 72 control words into categories which they deemed appropriate. Only 12 of the 576 experimental words sorted were misplaced, indicating high agreement as to the appropriate class for each word.

A trial was defined as the presentation of 24 S-R pairs. The same eight S-R pairs were repeated three times within a trial for the specific groups. For the class groups the same eight responses were repeated three times, but were paired with a different stimulus word each time they occurred within a trial.

The word and number pairs were presented to *Ss* in a typical paired-associates manner in a Patterson-type memory drum at a 2:2-sec. rate with a 24-sec. intertrial interval. Two different presentation orders

were used. The *Ss* responded into a microphone which was part of a voice key used to obtain response latencies in addition to the usual correct response measure. Before beginning L-I all *Ss* were informed that the responses would be the numbers one through eight, and were shown a list of the 72 words grouped into the 16 classes, but *S* did not know which number would be paired with which word, or class of words. One week later *Ss* returned for the recall trial, and L-II.

*Subjects.*—The 21 *Ss* in each group were elementary psychology students of both sexes fulfilling their subject-hour requirement at the Ohio State University.

Since responding to classes of stimuli was suspected to be more difficult than responding to specific stimuli, two experiments were conducted. In Exp. I the original learning was to a criterion of one perfect trial. In Exp. II, which employed 84 additional *Ss* from the same population, L-I was presented for five trials only. The experiments were identical in other respects. Thirty-five additional *Ss* completed L-I, but failed to return for the second session.

In summary, *Ss* learned the same number of responses to specific words, or to classes of words. At a recall trial no differences between their scores were expected. A transfer task served to determine whether *Ss* in the class group were responding to the stimuli as members of classes, and not as specific words.

## Results

The mean number of correct responses for all groups is plotted as a function of successive blocks of eight responses for each condition in Fig. 1. Although a trial was defined as 24 responses, there were only eight different responses used in the experiments. Therefore, each trial is a composite of three subtrials; it is these subtrials which are plotted on the abscissa in Fig. 1. Unless noted otherwise, all the statistical analyses were computed for a total trial of 24 responses, however. The *F* ratios computed for all four groups were based on 3 and 80 *df*. The *F* ratios for individual degrees of freedom were based on 1 and 80 *df*.

Table 2 presents the trials to cri-

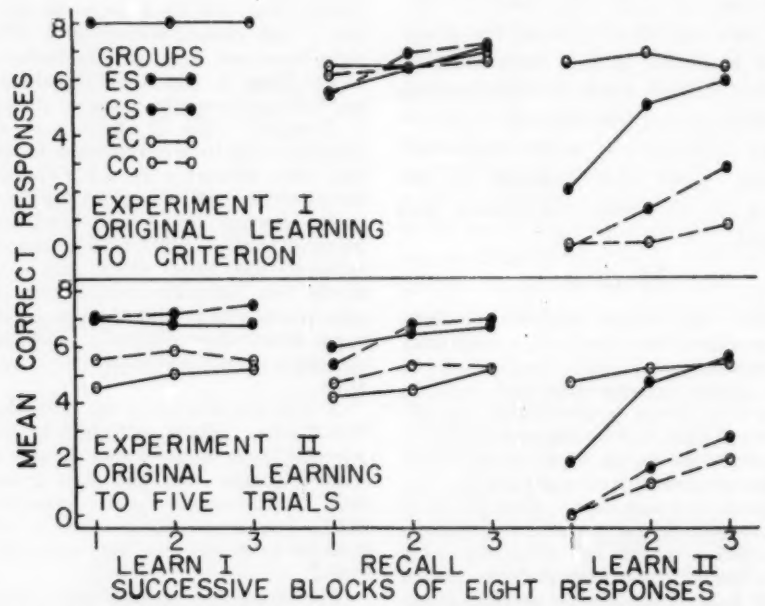


FIG. 1. Mean correct responses for successive blocks of eight responses. Each point on the abscissa represents one subtrial containing all eight responses from a total trial of 24 responses. The last three subtrials of the original learning are plotted at Learn I. All three recall subtrials are plotted at Recall. The first three subtrials of the transfer task are plotted at Learn II.

TABLE 2

TRIALS TO CRITERION AND RECALL IN EXP. I

| Group | L-I Trials | | Recall | |
|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* |
| ES | 6.81 | 2.67 | 19.19 | 3.02 |
| CS | 7.05 | 2.05 | 19.71 | 4.07 |
| EC | 9.14 | 2.04 | 20.10 | 4.70 |
| CC | 8.86 | 2.51 | 19.43 | 4.72 |

terion and recall scores for Exp. I. The $F$ ratio for trials is significant ($F = 5.33$, $P < .01$). Snedecor's (1946) test for individual degrees of freedom indicates that the class groups require more trials to criterion than do the specific groups ($F = 15.76$, $P < .01$). At recall the between-groups variance estimate is less than the within-groups variance, indicating no differences in recall among the groups. A second analysis of variance for the first eight responses yielded another insignificant $F$ ratio.

Table 3 contains the correct responses at L-I (Trial 5) and at recall for Exp. II. Both $F$ ratios are significant ($F = 12.54$, $P < .01$; $F = 7.17$, $P < .01$). By Trial 5 of L-I the specific groups have made more correct responses than the class groups ($F = 33.97$, $P < .01$). At recall the specific groups maintain this superiority ($F = 20.25$, $P < .01$). Neither class nor specific groups differ significantly from each other in any of the above comparisons.

Since the class and specific groups differed in their rates of learning, computation of loss scores (Richardson & Underwood, 1957; Underwood, 1954) provides a more appropriate assessment of retention than do the raw recall scores. Loss scores serve to equate associative strength at the end of learning. They are obtained by subtracting raw recall scores from estimated immediate recall scores

TABLE 3

CORRECT RESPONSES AT L-I AND RECALL IN EXP. II

| Group | Trial 5, L-I | | Recall | |
|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* |
| ES | 20.62 | 3.85 | 19.27 | 4.29 |
| CS | 21.86 | 3.13 | 19.19 | 4.70 |
| EC | 14.86 | 4.45 | 13.86 | 4.03 |
| CC | 17.00 | 4.68 | 15.48 | 5.10 |

which are derived from successive probability curves for each group. In order to obtain an estimated immediate recall score for 24 responses three successive probability curves based on eight responses were computed for each group. The first of these curves indicated the probability of a correct response on the next subtrial following a given number of prior correct responses. The second curve indicated the probability of a correct response on the second subtrial following a given number of prior correct responses, and the third curve provided probabilities for the third subtrial.

Table 4 contains the mean loss scores. Bartlett's test indicated homogeneity of variance in both experiments ($\chi^2 = 4.21$ and $1.46$; $df = 3$). Analysis of variance resulted in $F$ ratios of less than one in both experiments. Loss scores were also computed for the first subtrial at recall; again the $F$ ratios in both ex-

TABLE 4

MEAN LOSS SCORES

| Group | Exp. I | | Exp. II | |
|---|---|---|---|---|
| | Mean | $\sigma_M$ | Mean | $\sigma_M$ |
| ES | 3.95 | 0.69 | 2.47 | 0.71 |
| CS | 3.47 | 0.93 | 3.36 | 0.72 |
| EC | 2.96 | 1.06 | 2.95 | 0.89 |
| CC | 3.76 | 1.03 | 3.47 | 0.75 |

periments were not significant. The groups did not differ in loss scores in either experiment.

When *Ss* do not learn to a criterion, as in Exp. II, some items receive zero reinforcements. The loss scores in Table 4 were obtained by assigning zero probability to such items. Since this method may have underestimated loss scores, a second set of loss scores was obtained by assigning items with zero reinforcements the probability following one reinforcement. This procedure overestimated loss scores. Analysis of variance for the overestimated loss scores was also insignificant ($F = 2.13$, $P > .05$). This $F$ ratio is larger than the $F$ obtained using the first method due to increased loss for the class groups which had more items with zero reinforcements.

The transfer data (L-II, Trial 1) are shown in Table 5. The $F$ ratio in Exp. I was significant ($F = 109.78$, $P < .01$). The test for individual degrees of freedom indicated that the experimental groups differed from the control groups ($F = 285.38$, $P < .01$), that Group EC exceeded ES ($F = 36.41$, $P < .01$), and that Group CS exceeded CC ($F = 7.56$, $P < .01$). In Exp. II the $F$ of 46.31 was again significant ($P < .01$). The rank order of the groups was identical to Exp. I, but the differences were not as great; the experimental groups differed at the 5% level, and the control groups did not differ.

TABLE 5

CORRECT RESPONSE AT L-II, TRIAL 1, IN BOTH EXPERIMENTS

| Group | Exp. I | | Exp. II | |
|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* |
| ES | 13.00 | 4.38 | 12.24 | 5.07 |
| CS | 4.24 | 3.19 | 4.57 | 2.56 |
| EC | 20.00 | 4.78 | 15.29 | 4.31 |
| CC | 1.05 | 1.29 | 3.29 | 2.83 |

The latency data paralleled the results for correct responses and are not presented. They indicated that responding to classes of stimuli required more time than responding to specific stimuli. However, at L-II in both experiments the latency of Group EC was as short as that of the specific groups. The *Ss* in Group EC responded to new instances just as quickly as *Ss* in Group ES responded to new specific stimuli.

## DISCUSSION

The results of Exp. I indicated no differences in recall when the number of responses was equal for both class and specific groups, and when associative strength at the end of learning was equated. The three successive probability curves for all groups coincided after, at most, eight reinforcements. In the class groups 6% of the items had less than eight reinforcements, compared with 27% in the specific groups. The estimated immediate recall for the class groups was thus largely based on probabilities that did not differ from the probabilities of the specific groups; the same relationship holds for three-fourths of the probabilities contributing to the estimated immediate recall of the specific groups. In addition, the remaining one-fourth of the probabilities for the specific groups, i.e., those based on less than eight reinforcements, were higher than those for the class groups. The over-all effect is to produce equivalent estimates of immediate recall for all groups. When neither immediate recall nor raw recall scores differ, then, of course, the loss scores will not differ.

In Exp. II the class groups learned fewer items than the specific groups in five trials; the former groups also recalled fewer items. The three successive probability curves for all groups again coincided after, at most, eight reinforcements. In contrast to Exp. I, *Ss* in the class groups received less than eight

reinforcements for 71% of the items compared to 45% for the specific groups. The estimated immediate recall for the class groups was, therefore, largely derived from those parts of the probability curves where the probabilities were lowest. This resulted in lower immediate recall scores relative to the specific groups. The raw recall scores for the class groups were also lower. Again the over-all effect was to produce equivalent loss scores for the class and specific groups.

The transfer data from both experiments strongly suggest that the class groups were responding to the words as if they were members of a class. At the first presentation of a second list containing new instances from the old classes, Group EC was superior to Group ES. It would have been possible for *Ss* in Group ES to classify the stimuli during L-I since they had responded to these words many times before coming to the laboratory, but if they had done so, their performance during the first trial of L-II should have equaled that of Group EC. Under such conditions the class groups would ordinarily be predicted to display superior retention; they did not. When the number of responses was held constant and when associative strength at the end of learning was equated, retention did not differ between *Ss* who responded to stimuli as members of classes and *Ss* who responded to specific stimuli.

## Summary

No differences in retention between *Ss* responding to classes of stimuli, and *Ss* responding to specific stimuli were predicted if the number of responses in both conditions was equal. A transfer task was used to test the expectation that *Ss* would respond to the stimuli as instances of classes if they were given an opportunity. When original learning was carried to a criterion specific groups learned at a faster rate than class groups. There were no differences in either raw recall scores or loss scores between the groups. When original learning was for five trials the specific groups made more correct responses, and their raw recall scores were greater than the class groups. Loss scores after incomplete original learning did not differ. The greater transfer effects for the class groups supported the notion that they were responding to the stimulus words as members of classes. Under these conditions retention of responses learned to stimulus classes did not differ from retention of responses to specific stimuli.

## References

EBBINGHAUS, H. *Memory.* Trans. by H. A. Ruger and C. E. Bussenius. New York: Teachers Coll., Columbia Univer., 1913.

KATONA, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.

REED, H. B. Factors influencing the learning and retention of concepts: I. The influence of set. *J. exp. Psychol.*, 1946, **36**, 71–87.

RICHARDSON, J., & UNDERWOOD, B. J. Comparing retention of verbal lists after different rates of acquisition. *J. gen. Psychol.*, 1957, **56**, 187–192.

SNEDECOR, C. W. *Statistical methods.* Ames: Iowa State Coll. Press, 1946.

THORNDIKE, E. L., & LORGE, I. *The teachers' word book of 30,000 words.* New York: Teachers Coll., Columbia Univer. Press, 1952.

UNDERWOOD, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, **51**, 276–282.

(Received February 21, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 1, 1960

# GROWTH OF SENSATION ON SEVEN CONTINUA AS MEASURED BY FORCE OF HANDGRIP

JOSEPH C. STEVENS, JOEL D. MACK, AND S. S. STEVENS

*Harvard University*

There is abundant evidence that on prothetic continua the subjective magnitude ($\psi$) grows approximately as a power function of the stimulus magnitude ($\phi$) and that each modality has its characteristic exponent (S. S. Stevens, 1957). The present study is another step in a program designed to show whether functions of this form are able to predict the behavior of observers (*O*s) whose task it is to match the apparent intensity of sensations aroused through different sense modalities. If two continua are governed by power functions,

$$\psi_1 = \phi_1^m$$

and

$$\psi_2 = \phi_2^n$$

it follows that cross-modality matches, in which $\psi_1$ is equated to $\psi_2$, will determine an equal-sensation function of the form,

$$\phi_1 = \phi_2^{n/m}.$$

In log-log coordinates this equation determines a straight line with a slope equal to the ratio $n/m$.

Such equal-sensation functions have been both predicted and obtained by S. S. Stevens (1959a, 1959b) and by J. C. Stevens and Mack (1959). In the present study the forms of five more functions are predicted and to a good approximation verified. The *O*s were asked to produce, on a hand dynamometer, forces that seemed subjectively to match the apparent intensity of various stimulus levels produced by electric shock, white light, white noise, 1000-cycle tone, and vibration on the finger tip. In other words, instead of the usual procedure of asking *O* to estimate numerically the apparent loudness, say, of various stimulus levels, he was asked to express his appreciation of loudness by squeezing. Instead of emitting numbers, as it were, *O* emitted squeezes.

Since the subjective force of handgrip has been shown to grow as the 1.7 power of the physical force exerted (J. C. Stevens & Mack, 1959), the equal-sensation functions ought to have the form,

$$F = I^{n/1.7}$$

where $F$ is force and $I$ is intensity on the continuum of the criterion stimulus to which the apparent force is matched. The exponent $n$ is the exponent governing the growth of subjective magnitude as determined in earlier ratio-scaling experiments, mainly by the method of magnitude estimation. In Table 1 will be found the continua, the approximate value of the exponent ($n$) for each continuum, the predicted exponent of the equal-sensation function ($n/1.7$), and the exponent actually obtained by matching with handgrip.

## APPARATUS AND PROCEDURE

The *O* used his right hand to exert the matching forces on a hand dynamometer. This dynamometer (J. C. Stevens & Mack, 1959), permitted the measurement of force up

[1] This research was supported in part by Grant G2668 from the National Science Foundation, and in part by Contract N5onr-1866(15) with the Office of Naval Research (Project Nr142-201, Report PNR-226). Reproduction for any purpose of the United States Government is permitted.

to 100 lb. and was designed to align itself, by means of ball-bearing joints, in the direction of the force applied to its handle. For a full 100-lb. deflection of the dial, it was necessary to displace the handle through a distance of only .039 in.

In each experiment various intensities of a criterion stimulus were presented in a different irregular order to each *O*. The various stimulus levels were then repeated one or more times, again in a different order for each *O*. The *O* was instructed to squeeze the dynamometer with a force that seemed to match the apparent magnitude of the criterion stimulus. He was to do this by "bracketing" and was to signal as precisely as possible the moment at which the matching force was achieved.

Twenty-six different *O*s participated in one or more of the five experiments, 4 in four of the experiments, 3 in three of them, 8 in two of them, and 11 in a single experiment. Three of the *O*s were staff members of the laboratory and the rest were students in the University.

*Electric shock.*—The apparatus was essentially that used by Stevens, Carton, and Shickman (1958) to determine a ratio scale for the apparent intensity of electric shock. The *O* immersed the first and third fingers of his left hand (up to the first joint, approximately) in salt-water electrodes. A 60-cycle current was passed through these electrodes and through a series resistor of 250,000 ohms. The strength of the current through the fingers was determined by measuring the voltage drop across a 10,000-ohm resistor substituted for *O*'s fingers. The current was varied by means of a Variac transformer. A timer in the circuit limited the stimulus duration to .6 sec. at each presentation. The apparent intensities of five levels of current (.29, .41, .52, .62, and .72 milliamp.) were matched by having each of the 10 *O*s squeeze the dynamometer twice. Here, and in all the other experiments, the order of the presentation of the criterion stimuli was "random," except that neither of the two extreme values was ever presented first.

*White light.*—The target was a circular patch, of homogeneous luminance, subtending 5 degrees of arc when viewed from a distance of 17 in. The luminance of the patch was controlled by means of neutral density filters. The *O*, who was dark-adapted for about 10 min. before the session, looked through a binocular aperture at the end of a black hood enclosing the target. A key, pressed by *O*, operated a timer that lighted the source for 2 sec. There were six stimulus levels: 56, 66, 76, 86, 91, and 96 db re $10^{-10}$ Lambert. (In this notation, 70 db corresponds to 1 mL.). Each stimulus was matched twice by each of 10 *O*s.

*White noise.*—The output of a white noise generator was fed through an attenuator to a pair of PDR-8 earphones mounted in MX-41/AR cushions. The *O* turned on the stimulus by depressing a switch in the circuit. He was instructed to listen to the noise for 1–2 sec. at a time and to try to match the loudness by squeezing the dynamometer. Five levels were used as stimuli: 55, 65, 75, 85, and 95 db re .0002 dyne/cm.$^2$. Each was matched four times by each of 10 *O*s.

*1000-cycle tone.*—The apparatus was the same as that used in the experiment on white noise, except that the signal source was an audio oscillator set at 1000 cps. Six levels were used as stimuli: 47, 57, 67, 77, 82, and 87 db re .0002 dyne/cm.$^2$. Each was matched twice by each of 10 *O*s.

*Vibration.*—A 60-cycle signal from an audio oscillator was fed through an attenuator into a matching transformer whose output activated a V47 Goodmans Vibration Generator. The *O* rested his left middle finger tip lightly on a plastic button (about ¼ in. in diameter) affixed to the armature of the vibrator. He was instructed to turn on the stimulus by depressing a switch for a second or two at a time, and to try to match its intensity by squeezing the dynamometer. The six stimulus values were produced by impressing, across the terminals of the vibrator, voltages of 27, 37, 43, 49, 53, and 57 db re 1 mv. On this scale, threshold is in the vicinity of 10 db. Each stimulus was matched three times by each of 12 *O*s.

## Results

The median results for each of the five experiments are shown in Fig. 1, together with the results of the two earlier experiments (J. C. Stevens & Mack, 1959), in which *O*s endeavored to exert forces that matched the apparent heaviness of lifted weights and the apparent intensity of pressure applied to the passive skin of the palm. The points in Fig. 1 determine in log-log coordinates a set of functions that are approximately linear but widely different in slope. In other words, these equal-sensation functions approximate power func-

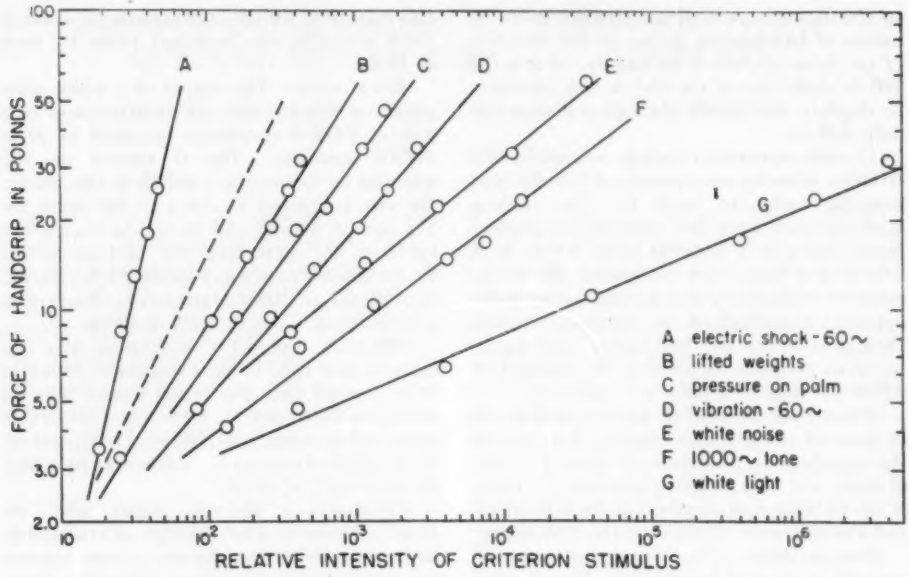


FIG. 1. Seven equal-sensation functions. The relative positioning of the functions along the abscissa is arbitrary. (For the absolute stimulus values, see text.) The dotted line shows a slope of 1.0 in these coordinates. Each point stands for the median force exerted to match a criterion stimulus.

tions having exponents of different sizes. The exponents (slopes) of the lines in Fig. 1 (determined by least squares) are listed in Table 1, where they may be compared with the exponents predicted from previous experiments in which the slopes were determined by "numerical" methods (e.g., magnitude estimation). The outcome of the handgrip experiments seems to confirm the form and the exponents of the several functions

TABLE 1

THE EXPONENTS (SLOPES) OF EQUAL-SENSATION FUNCTIONS, AS PREDICTED FROM RATIO SCALES OF SUBJECTIVE MAGNITUDE, AND AS OBTAINED BY MATCHING WITH FORCE OF HANDGRIP

| Continuum | Ratio Scale | | Exponents Determined by Handgrip | |
|---|---|---|---|---|
| | Exponent of Power Function | Reference | Predicted Exponent | Obtained Exponent |
| Electric shock (60-cycle current) | 3.5 | S. S. Stevens, Carton, & Shickman (1958) | 2.06 | 2.13 |
| Brightness of white light | 0.33 | S. S. Stevens & Galanter (1957) | 0.20 | 0.21 |
| Loudness of white noise (SPL) | 0.6 | J. C. Stevens (1958) | 0.35 | 0.41 |
| Loudness of 1000-cycle tone (SPL) | 0.6 | S. S. Stevens (1955, 1956) | 0.35 | 0.35 |
| 60-cycle vibration | 0.95 | S. S. Stevens (1959b) | 0.56 | 0.56 |
| Heaviness of lifted weights | 1.45 | S. S. Stevens & Galanter (1957) | 0.85 | 0.79 |
| Pressure on palm | 1.1 | J. C. Stevens & Mack (1959) | 0.65 | 0.67 |

determined independently by the ratio scaling methods based on numerical estimation.

*Variability.*—The variability of handgrip matching turns out to be about as great as the variability of numerical judgments of sensory magnitude. For a given continuum, the interquartile range of the forces exerted to match a criterion stimulus is approximated by the formula:

$$\log Q_3 - \log Q_1 = k$$

where $k$ has a different value for each continuum matched. From Table 2, which lists the median value of $k$ for each of the seven continua, it appears that it was harder to match some continua than others. The value of $k$ for a 1000-cycle tone, for example, is approximately 2.5 times greater than its value for skin pressure. Differences of this kind were also reflected in the comments of some *O*s about the difficulty of the task. It is an interesting fact that the continuum that has been most assiduously explored in this country and elsewhere, namely, the loudness of a 1000-cycle tone, is perhaps the most variable and difficult of these seven. At any rate, the value of $k$ seems to be roughly the same size as comparable measures of the variability of magnitude estimates (see S. S. Stevens, Carton, & Shickman, 1958; S. S. Stevens, 1956; J. C. Stevens, 1958).

TABLE 2

VARIABILITY OF THE MATCHES

$k$ is the median of the interquartile ranges (log $Q_3$ − log $Q_1$). In order to obtain $k'$, each *O*'s judgments (squeezes) were multiplied by a factor such that the judgments of all *O*s were reduced to a common modulus. $k'$ is the median interquartile range (log $Q_3$ − log $Q_1$) of the values obtained by this procedure.

| Continuum | $k$ | $k'$ |
|---|---|---|
| Electric shock | 0.35 | 0.20 |
| White light | 0.30 | 0.22 |
| White noise | 0.26 | 0.20 |
| 1000-cycle tone | 0.48 | 0.23 |
| Vibration | 0.36 | 0.18 |
| Lifted weights | 0.19 | 0.12 |
| Pressure on palm | 0.25 | 0.16 |

In the task of matching a continuum such as handgrip to each of various other continua, there are several sources of variability. In addition to instrumental errors and the inherent variability of human judgment, some of the variability is attributable to individual differences among *O*s. These latter differences seem to manifest themselves in two principal ways: as a difference in the slope and (or) as a difference in the position (intercept) of the equal-sensation function.

Although differences in slope may arise from several causes, they sometimes reflect genuine differences in the action of sensory transducers, e.g., recruitment in the ear or in the eye (S. S. Stevens, 1958). Actually, whether the slope is diagnostic of the operating characteristic of the sensory system depends greatly on what is matched to what. Thus differences in slope obtained with methods that involve numerical estimation usually arise from the simple fact of human variability. On the other hand, comparable differences in slope obtained by matching, say, loudness in one ear to loudness in the other, would generally indicate fundamental differences in the two loudness functions.

As regards the position of an equal-sensation function, differences here nearly always reflect differences in the judgmental criterion. That is to say, the absolute value of the force of handgrip exerted to produce a match may vary considerably from one *O* to another, for the simple reason that a "perfect" match is impossible and *O* must choose some criterion of what he will call equal. This effect is analogous to the use of a different modulus by each *O* in those magnitude-estimation experiments in which

no standard stimulus is presented (S. S. Stevens, 1956). A major component of this variability can be removed by multiplying the results of each *O* by whatever factor is needed to bring the results into coincidence with those of the rest of the *O*s at some selected point.

This procedure of adjusting the judgments (squeezes) to a common "modulus," so to speak, was tried for each of the seven continua. The median results are shown in Table 2 under the heading $k'$. It is seen that the total variability is considerably reduced when the "modulus" factor is eliminated.

*Inverse matchings.*—The hand dynamometer lends itself reasonably well to the procedure under which *O* produces squeezes to match the apparent intensities of other sensations. It is adapted less well to the inverse task in which *O* adjusts the level of some other sensation to match the apparent force of a given squeeze. For this inverse procedure *O* is required to squeeze until *E* signals that the proper pull has been reached, and then *O* proceeds to adjust another stimulus (e.g., sound intensity) until it seems to match the force he has exerted. Exploratory experiments were conducted with this procedure for two types of stimuli: white noise and 60-cycle vibration.

The results for the loudness of white noise were quite satisfactory, in the sense that the slope of the equal-sensation function was close to that shown by Line *E* in Fig. 1. As in other cross-modality matchings (S. S. Stevens, 1959a), there appeared a small "regression" or "centering" effect which made the function slightly steeper when the loudness was adjusted than when the force of handgrip was adjusted.

The results for vibration, however, were quite unsatisfactory, for the slope of the equal-sensation function was much steeper than expected. This large discrepancy is due mainly to the fact that *O*s do not have three hands. Rather than devise some novel method by which *O* could adjust the vibration, he was asked to squeeze the required amount with his right hand, and then to reach over and adjust, again with his right hand, the vibration amplitude applied to the middle finger of his left hand. This was apparently too much to ask, especially when matchings were required for a wide range of criterion forces. For six forces ranging from 4 to 37 lb., the function obtained had a slope of about 0.94 (compared with 0.56 for Line *D* in Fig. 1). This value dropped to about 0.73 when the range was made 5 to 25 lb. In a third experiment in which only three criterion forces (7, 12, and 20 lb.) were used, the slope was about .80. Needless to say, the variability of the vibration adjustments was large. In all three experiments the over-all range of the settings produced to match a given force was consistently between 20 and 30 db. It appears that *O* cannot "remember" a force well enough to carry out the delayed adjustment of a matching vibration with reasonable precision. Under these conditions the magnitude of the "regression" effect becomes extremely large.

Presumably a successful matching of vibration to handgrip might be possible if *O* could vary the amplitude of the vibration by means of some type of foot control. This would permit more or less simultaneous comparison of the two sensations.

## Discussion

One way of regarding the results presented in Fig. 1 is to see them as a possible reassurance to those who express misgivings concerning the use of numerical estimation in the measurement of sensory effects. In obtaining the results with the dynamometer, we have effectively avoided any demand on "number behavior," and we have not called upon any explicit concepts of ratio or proportion. All that was asked was that *O* try to assess the magnitude of a given sensation by producing another sensation of comparable intensity. The equating of the apparent strengths of two different sensations is a procedure sanctioned by much usage, as in the determination of equal loudness contours, heterochromatic brightness matches, etc., and presumably few people object to it on principle. In the present instance, the traditional procedure has been extended to the matching of apparent intensities across two modalities. Depending on the modalities involved, this task may be harder or easier than, for example, the matching of the loudnesses of two sounds having disparate qualities.

Granted that squeezes can be matched to other sensations, the hand dynamometer provides a method for measuring sensory intensity in the sense of determining the *relative* form of the over-all *operating characteristic* of the sensory system. Thus the data in Fig. 1 show that the sensation of electric shock grows with an acceleration about six times greater than loudness. In other words, the relative slopes of the functions in Fig. 1 have a significant meaning, for they describe the relative rates of growth of the various sensations. Furthermore, if we were willing to take any one of these sensory continua as a standard of reference, we could proceed to use it as a yardstick with which to measure sensory intensity on each of the other continua. This procedure would be useful for certain purposes, but the scales thereby achieved would have a basic limitation. Although they would presumably be power functions of the *linear* ratio scale of sensory intensity, the degree (exponent) of the power function would not be known.

What we would like, of course, would be to know the forms of the several functions relative to that of a linear continuum. But there seems to be in this world no *absolute* criterion of linearity; if everything in the universe were curved how would we know it? Hence the choice of a reference continuum is to some extent arbitrary. On the other hand, as a plausible reference continuum (the one assumed to be linear), the number continuum itself, when used by a person with proper understanding of it, has much to recommend it. The number system is readily accessible, and it "agrees" quite well with certain other continua (e.g., apparent visual length) on which linearity of subjective magnitude may reasonably be expected. What this means concretely is that, when *O*s are asked to assign numbers to apparent lengths of lines, the result is usually a power function with a slope close to 1.0. Thus a 4-in. line looks about twice as long as a 2-in. line.

On the assumption, then, that *O*'s number scale may be regarded as linear, it becomes a matter of interest to compare two kinds of experiments, one in which *O* emits squeezes to indicate magnitude of sensation and one in which he emits numbers for the same purpose. This comparison can be made in two stages. One question concerns the *relative* values of the exponents (slopes in log-log coordinates) obtained in the two experiments, and the other question concerns the *absolute* values of these exponents.

It is clear from Fig. 1 and Table 1 that the rank order of the slopes (exponents) is the same whether *O* matches the various sensory continua with squeezes or with numbers. Furthermore, to a good approximation, the slopes obtained by "handgrip matching" are a constant submultiple (1.7) of the slopes obtained by "number matching." To the extent that this proportionality obtains, we have a method by which the "relative" exponents obtained by squeezing can be converted into "absolute" exponents.

In the present instance the factor of proportionality (1.7) was derived from previous experiments designed to match, so to speak, number continuum against apparent force of handgrip. This kind of matching exhibits the usual "regression" effect referred to above, in the sense that when *O* adjusts handgrip to match numbers (magnitude production) the slope is steeper than when he adjusts numbers to match handgrip (magnitude estimation). This raises an interesting theoretical point. Which, if either, of these two slopes can be expected to correspond to the factor of proportionality? To the extent that both these slopes are biased by the "regression" effect, it seems that neither would correspond in general to the factor of proportionality. On the other hand, if it happens that the effects on the slope produced by interchanging standard and variable stimulus tend to be equal and opposite, then a slope intermediate between the two obtained slopes should provide the desired factor.

This, indeed, is what seems to have worked out. The value 1.7 for the exponent of the function relating number

and handgrip was determined by pooling the results of magnitude production and magnitude estimation. With this relatively unbiased measure of the exponent determined, it becomes possible to substitute squeezes for numerical estimation, and then to recover the results characteristic of numerical estimation merely by multiplying the "squeeze" exponents by 1.7. The entries in Table 1 show how well this has worked out.

Among the advantages of validating the measurement of sensory intensity by methods that do not require *O* to use numbers, are the possibilities opened up for the diagnosis and study of idiosyncratic tendencies in numerical estimations. When a person's numerical estimates of a variable-like loudness produce a function that differs consistently and systematically from the median function, it is always an interesting question whether his hearing is different or whether he simply estimates differently. If his loudness function does not differ consistently from "normal" when determined by several cross-modality matchings, e.g., by squeezing, by matching loudness to vibration, etc., it would suggest that the dynamic operating characteristic of this person's auditory system is probably not peculiar. On the other hand, when the slope of the loudness function for a particular individual maintains a given peculiarity under several cross-modality matchings, evidence may be at hand that the peculiarity is real. In this fashion it would, in principle, be possible to diagnose such clinically important stigmata as recruitment, or its absence, in a hard-of-hearing person. Since the indicated therapy may be quite different in the two cases, this is one of those instances in which a valid measurement of the growth function governing a sensation is a matter of grave concern.

A final comment is in order regarding a theory that is alleged to account for the form of the various subjective scales. This theory revives an old notion about judgment, to the effect that estimations of sensory magnitude are based on experience with some physical measure of stimulus magnitude, i.e., the judge commits the "stimulus error." Warren (1958) elaborates on this idea and produces what he calls the *physical-correlate theory* of sensory intensity. This theory "considers that judgments of sensory magnitudes are based upon experience with the manner in which sensory excitation is correlated with the amount of some physical attribute associated with the stimulus" (Warren, 1958, p. 676). In his search for a plausible basis on which people might learn to judge such things as loudness and brightness, Warren comes up with the hypothesis that judgments of these continua are based on our experience with the inverse square law. (Actually the inverse square law does not predict correctly the exponents of the sone and bril scales, but Warren undertakes to explain away this defect by suggesting that the sone and bril functions were determined under "unfamiliar" conditions.)

A "theory" of this sort, since it is probably not susceptible to proof or disproof, must be judged, if at all, on the basis of the demands it makes upon our credulity. Some will find it difficult to believe that the relations among the exponents of all the functions in Fig. 1 are explainable on the basis of *O*'s experience with the physical stimuli involved. Where and how, for example, did *O*s "learn" the function for electric shock, or for vibration? And is it plausible to attribute to our prior experience with "physical correlates" the fact that the relations among the sensory scales remain invariant under a wide variety of cross-modality matchings?

Another objection to this theory is that it is essentially negative. If taken seriously, it would seem to imply that neither magnitude estimations nor cross-modality matchings are able to tell us anything about the operating characteristics of a sensory system. Pushed to an extreme, it would deny that we could tell the difference between a normal ear and an ear with recruitment, for if all sensory judgments are conditioned only by prior

learning, the properties of the end organ and its associated neural structure cannot matter. Still, it is easy to change the form of a sensory function by altering the state of an end organ, for example, by light-adapting the eye.

Since the theory is consistent with the power-function relation between stimulus and response, it is claimed that "the physical-correlate theory provides a basis for Stevens' empirical law" (Warren, Sersen, & Pores, 1958, p. 707). Although it would indeed be nice to find a general basis for this psychophysical law, the law itself would seem to lose much of its interest if it turned out that the characteristics of the sensory transducers have nothing to do with the exponents of the various power functions. It is not, of course, unreasonable to suppose that experience with physical correlates is an important factor in determining numerical judgments on some kinds of perceptual continua (e.g., length of lines), and it is certainly true that an arbitrary scale can be learned (as when acoustical engineers learn to estimate sound levels in decibels). Nevertheless, it is not unreasonable to believe that the slopes of functions like those in Fig. 1 may depend more on the nature of the organism and its sensory transducers than on experience with such "physical correlates" as the inverse square law.

## Summary

The *Os* produced forces on a hand dynamometer to match various levels of subjective intensity for five kinds of criterion stimuli: electric current applied to the fingers, white light, white noise, a 1000-cycle tone, and a 60-cycle vibration applied to the finger tip. It is shown that the forms of the five equal-sensation functions (plus two functions obtained earlier for lifted weights and for pressure applied to the palm) can be predicted by the ratio scales of subjective intensity determined independently by the procedure of numerical estimation. The results add further evidence to the validity of the general psychophysical law that subjective magnitude grows as a power function of stimulus magnitude. Although force of handgrip cannot by itself be used to determine the absolute values of the power-function exponents, it can be used to determine the *relative* values of the exponents for a wide variety of prothetic continua. Taken in conjunction with the previously determined scale for the subjective magnitude of handgrip (exponent = 1.7), the matching of handgrip to various stimulus intensities confirms the approximate *absolute* values of the exponents for seven different continua.

## References

Stevens, J. C. Stimulus spacing and the judgment of loudness. *J. exp. Psychol.*, 1958, **56**, 246–250.

Stevens, J. C., & Mack, J. D. Scales of apparent force. *J. exp. Psychol.*, **58**, 405–413.

Stevens, S. S. The measurement of loudness. *J. Acoust. Soc. Amer.*, 1955, **27**, 815–829.

Stevens, S. S. The direct estimation of sensory magnitudes—loudness. *Amer. J. Psychol.*, 1956, **69**, 1–25.

Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, **64**, 153–181.

Stevens, S. S. Some similarities between hearing and seeing. *Laryngoscope*, 1958, **68**, 508–527.

Stevens, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *J. exp. Psychol.*, 1959, **57**, 201–209. (a)

Stevens, S. S. Tactile vibration: Dynamics of sensory intensity. *J. exp. Psychol.*, 1959, **57**, 210–218. (b)

Stevens, S. S., Carton, A. S., & Shickman, G. M. A scale of apparent intensity of electric shock. *J. exp. Psychol.*, 1958, **56**, 328–334.

Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, **54**, 377–411.

Warren, R. M. A basis for judgment of sensory intensity. *Amer. J. Psychol.*, 1958, **71**, 675–687.

Warren, R. M., Sersen, E. A., & Pores, E. B. A basis for loudness-judgments. *Amer. J. Psychol.*, 1958, **71**, 700–709.

(Received February 24, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 1, 1960

# THE RELATIONSHIP BETWEEN DIRECTION AND AMOUNT OF STIMULUS CHANGE AND AMOUNT OF PERCEPTUAL DISPARITY RESPONSE [1]

H. D. KIMMEL

*University of Florida*

In a recent experiment (Grings & Kimmel, 1958) a difference in magnitude of GSR in favor of unexpected over expected stimuli was found on test trials following a training series in which *S*'s expectations had been formed by confirming experiences. This difference in magnitude of response was referred to as a "perceptual disparity response" (hereafter called PDR), the name having been adapted from the work of Bruner (1951). The training series in that experiment involved following *S*'s right or left lever movements by the presentation of a light or tone, the pairing of right-light, left-tone, right-tone, or left-light being maintained consistently throughout the training series. On test trials either the trained (expected) or the other (unexpected) stimulus was presented after a particular lever movement. The *S* was free to move the lever to either side, both during training and test, except that he was instructed to avoid marked favoring of one side over the other. The PDR was at-attributed to a discrepancy between *S*'s expectation and what actually was presented.

In the present experiment the amount of discrepancy between expected and actual stimulation was varied along the quantitative stimulus dimension of physical intensity. The purpose of the study was to determine whether or not the amount of PDR is a function of the amount and/or the direction of the discrepancy between expected and actual stimulation. The stimulus dimension used to evaluate these questions was that of auditory intensity, defined in physical terms (i.e., decibels re .0002 dynes per $cm^2$). The data were gathered during the second part of an experiment on classical conditioning of the GSR, in which CS intensities had been varied along a physical continuum (Kimmel: 1958, 1959). Discrepancies between expected and actual stimulation were introduced at the beginning of an extinction series by changing the intensity that had been used during the acquisition series. It was assumed that *S* became set to receive the CS intensity experienced during acquisition, so that when the intensity was changed for the extinction series a discrepancy between expectation and actuality occurred. Sets and PDRs of this type have been demonstrated in previous research of this type (Grings, Dossett, & Honnard, 1959; Grings & Kimmel, 1958; Grings & O'Donnell, 1956).

Amount of PDR was specified for this study in terms of GSR magnitudes on the first trial after the change in stimulus intensity was introduced. The effect of stimulus intensity, per se, on amount of GSR (Hovland & Riesen, 1940) was controlled by

[1] Based on data collected as part of a dissertation submitted to the Graduate School, University of Southern California, 1958.

correcting test response magnitudes for stimulus intensity differences in a manner described below.

## METHOD

*Subjects and design.*—Ninety-six undergraduate students at the University of Southern California, 56 men and 40 women, served as volunteer *Ss*.[2] Each *S* was paid $1.00 for participating in the experiment.[3] Volunteers with more than 15 db hearing loss in either ear were excused from participating. Two volunteers were rejected on this basis.

The *Ss* were assigned randomly to 15 experimental groups according to the plan given in Table 1.[4] The groups differed according to the combination of acquisition and extinction stimulus intensities they received. As is indicated in Table 1, 13 different amounts of change in stimulus intensity were generated by this design, ranging from a −70 db change for the group shifted from an acquisition intensity of 115 db to an extinction intensity of 45 db, through zero change for the group receiving acquisition and extinction intensities of 75 db, to a +70 db change for the group shifted from an acquisition intensity of 35 db to an extinction intensity of 105 db. The groups shifted from acquisition intensities of 35 and 95 db to extinction intensities of 45 and 105, respectively, both experienced a +10 db change, while the groups shifted from acquisition intensities of 55 and 115 db to extinction intensities of 45 and 105 db, respectively, both experienced a −10 db change. Thus, the 15 combinations of acquisition and extinction intensities led to 13 amounts of change.

[2] Sixteen *Ss* were run by R. Honnard.

[3] Payments to *Ss* came from a grant from the National Science Foundation to W. W. Grings.

[4] The purpose of the first part of the study was to evaluate the effect of CS intensity on amount of conditioning during acquisition and extinction. A 5 × 3 design, replicated four times for *Ss* receiving paired tones and shocks, was generated with this purpose in mind. A 3 × 3 design, replicated four times for *Ss* receiving unpaired tones and shocks, was added to provide controls for sensitization effects in the paired group. Since the study reported here was a secondary consideration at the time of data collection, the *N*s in each of the 15 groups are not the same, nor are they the same in the 13 groups receiving different amounts of intensity change.

TABLE 1

NUMBERS OF *Ss* AND AMOUNTS OF INTENSITY CHANGE (ΔI) ASSOCIATED WITH 15 COMBINATIONS OF CS AND EXTINCTION STIMULUS INTENSITIES (+ INDICATES INCREASE IN INTENSITY; − INDICATES DECREASE IN INTENSITY)

| CS Intensity | Extinction Stimulus Intensity | | | | | |
|---|---|---|---|---|---|---|
| | 45 db | | 75 db | | 105 db | |
| | *N* | ΔI | *N* | ΔI | *N* | ΔI |
| 35 db | 8 | +10 | 8 | +40 | 8 | +70 |
| 55 db | 4 | −10 | 4 | +20 | 4 | +50 |
| 75 db | 8 | −30 | 8 | 0 | 8 | +30 |
| 95 db | 4 | −50 | 4 | −20 | 4 | +10 |
| 115 db | 8 | −70 | 8 | −40 | 8 | −10 |

*Apparatus.*—The equipment used in this study has been described in detail elsewhere (Kimmel: 1958, 1959). A Hewlett-Packard Model 200-B audio oscillator was used to generate a 1,000 cps tone. The duration of the tone was 4 sec., controlled electronically. The tone was delivered to *S*'s PDR-8 headphones (with MX-41/AR cushions) through a Hewlett-Packard Model 350-B decibel attenuator. Intensity was rated at *S*'s headphones (in db re .0002 dynes/cm.$^2$) and was varied by changing the setting of the attenuator.

The GSR was picked up by a saline paste-metal-skin junction using ¾-in. Lucite cups attached by rubber straps to the palm and back of *S*'s nonpreferred hand. Alcohol was used to clean the skin before the electrodes were attached. The response was measured as a resistance change in a Wheatstone bridge input to a Brush DC amplifier, with changes in resistance recorded as deflections on an Esterline-Angus ink-writing milliammeter with a paper speed of 3 in. per min. The current through *S* was constant at 42.5 microamperes.

*Instructions.*—Instructions were administered in writing. They informed *S* that this was an experiment on the way in which different stimuli evoke the GSR. No task was given to *S* other than attending to the stimuli, remaining as motionless as possible, and relaxing his body to prevent discomfort.

*Procedure.*—Observations were made in a soundproof room with all lights out except for a covered 7.5-w. lamp in *E*'s cubicle. A partition separated *S* from *E* and the equipment. The only noise in the room was the click of the signal magnets on the recorder

and the hum of an electric fan masking these clicks.

After receiving a hearing test and reading the instructions, *S* was blindfolded and the equipment attached. The *S* then received three presentations of the CS intensity and three of the extinction stimulus intensity (both being 1,000 cps tones of 4 sec. duration) in a counterbalanced order fixed for all *S*s. This preliminary presentation series was followed by three shocks to *S*'s preferred forearm (the shock intensity was set by *S* at the beginning of the session so that it would be mildly annoying) and six additional presentations of the CS intensity. These trials were all separated by 20–30 sec., varied unsystematically. Sixty *S*s then received 20 trials, separated by 40–60 sec., of the CS paired with shock (these 60 *S*s were divided into 15 groups of 4, each group receiving a different one of the 15 combinations of acquisition and extinction stimulus intensities), while the remaining 36 *S*s received a random sequence of unpaired presentations of 20 shocks and 20 tones, separated by 20–30 sec., varied unsystematically (these 36 *S*s were divided into 9 subgroups of 4, each group receiving a different one of the 9 combinations of acquisition and extinction intensities excluding those involving the 55 and 95 db acquisition intensities).

All 96 *S*s then received six presentations of the extinction stimulus without shock. The first of these extinction presentations provided the dependent measures for this report.

## Results

Magnitude of GSR was measured as the maximum change in skin resistance occurring within a 5-sec. interval beginning with the onset of the test stimulus. This measure was transformed into units of square root of conductance change. The transformed responses obtained on the three preliminary presentations (i.e., before acquisition) of the extinction stimulus were averaged separately for each *S* and this average was subtracted from *S*'s transformed response on the first extinction trial to provide a measure of PDR corrected for the effect of stimulus intensity differences, and, to some extent, for differences among *S*s in GSR responsiveness.

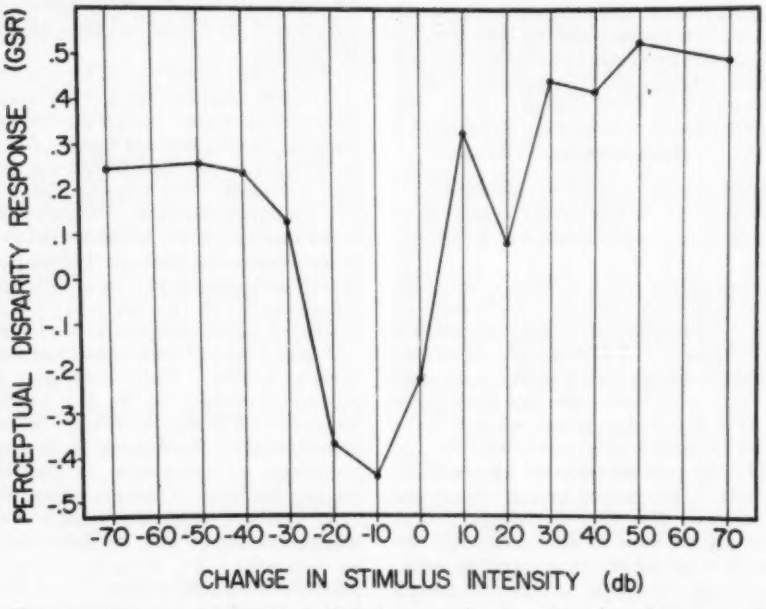


Fig. 1. Amount of perceptual disparity response as a function of amount of change in stimulus intensity.

Figure 1 is a graphical representation of the differences among mean PDRs associated with the 13 amounts of stimulus intensity change. The figure indicates that greater mean PDRs resulted from large amounts of intensity change than from small amounts of intensity change, regardless of the direction of the intensity change, and that greater mean PDRs resulted from increases in intensity than from decreases in intensity.

To simplify the statistical analysis of these data, *Ss* were divided into four subgroups according to whether they had received positive or negative intensity changes and whether they had received small or large amounts of intensity change. Thus 24 *Ss* who had received positive changes of 10, 20, or 30 db constituted one subgroup, 24 *Ss* who had received negative changes of 10, 20, or 30 db formed a second subgroup, 20 *Ss* who had received positive changes of 40, 50, or 70 db a third subgroup, and 20 *Ss* who had received negative changes of 40, 50, or 70 db a fourth subgroup. The 8 *Ss* who received zero change in intensity were omitted. Table 2 presents the means and *SDs* of PDRs for these four subgroups and illustrates the logic of the analysis of variance of PDRs presented in Table 3.

TABLE 2

PDR MEANS AND *SDs* ON THE FIRST EXTINCTION TRIAL

| Amount of Change in Intensity | | Direction of Change in Intensity | | Totals |
|---|---|---|---|---|
| | | Upward | Downward | |
| Small (10, 20, 30 db) | *M* | 0.32 | −0.24 | 0.04 |
| | *SD* | 2.44 | 1.89 | |
| | *N* | 24 | 24 | |
| Large (40, 50, 70 db) | *M* | 0.47 | 0.24 | 0.36 |
| | *SD* | 2.66 | 2.36 | |
| | *N* | 20 | 20 | |
| Totals | *M* | 0.39 | −0.02 | 0.18 |

TABLE 3

ANALYSIS OF VARIANCE OF PDRs ON THE FIRST EXTINCTION TRIAL

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Amount of intensity change | 1 | 2.146 | 5.38* |
| Direction of intensity change | 1 | 3.752 | 9.40** |
| Amount × Direction | 1 | 0.623 | |
| Error (within cells) | 84 | 0.399 | |

* $P = .05$.
** $P = .01$.

As is shown in Table 2 and evaluated statistically in Table 3, the mean PDR for large changes in intensity was significantly greater than the mean PDR for small changes in intensity ($F = 5.38$, 1 and 84 *df*) and the mean PDR for changes in an upward direction was very significantly greater than the mean PDR for changes in a downward direction ($F = 9.40$, 1 and 84 *df*). The absence of an interaction effect indicates that these differences occurred at both levels of both variables.

## DISCUSSION

The results of this study show quite clearly that the PDR observed by Grings and Kimmel (1958) is, at least in part, a function of the amount and direction of disparity between *S*'s expectation and actual stimulation, when stimulus intensity is the dimension along which the disparity occurs. In the present study a minor source of confounding was associated with the fact that the 35-db CS intensity was the only acquisition intensity that led to significant conditioning during the acquisition series (Kimmel: 1958, 1959). Since 12 *Ss* who had been conditioned with the 35-db CS all experienced increases in intensity, it is possible that the significant difference in mean PDR favoring upward over downward changes partly resulted from generalization of this conditioning effect. Two considerations argue against this possibility, however,

The first is the fact that 12 other *Ss* who had received the 35-db CS had not paired with shock and thus could not have been conditioned. Secondly, the differences in amount of conditioning that had been observed during the acquisition series had all but vanished before the end of the acquisition series and, thus, could not have carried over to the first extinction trial.

It is not surprising that PDR, or as it might be referred to in more general terminology, startle, was found to vary positively with the amount of difference between what *S* expects to receive and what he actually does receive. If startle results from receiving the unexpected, more startle should be found when the difference between expectation and actuality is great than when it is small. Of considerable significance, however, is the implication of the present findings that *S*'s expectation is sufficiently stimulus-bound, in the physical sense of the word "stimulus," that it is sensitive to differences in both amount and direction of intensity differences.

It would be of interest to determine whether the effects observed in this experiment can be duplicated using stimulus dimensions other than intensity and sense modes other than audition. Earlier studies (Grings, Dossett, & Honnard, 1959; Grings & Kimmel, 1958) found the effect when changes were made between sense modes and were, therefore, more qualitative than in the present study.

## Summary

A study was done, using 96 *Ss*, in which the intensity of an auditory stimulus (a 1,000-cps tone of 4 sec. duration) used during a conditioning series of 20 trials was changed for an extinction series. Thirteen different amounts of change in intensity were used, ranging from a 70 db downward change, through zero change, to a 70 db upward change. Amount of GSR, corrected for the effect of stimulus intensity differences, was used to measure the perceptual disparity response (PDR) occurring on the first trial with the changed intensity.

The results showed that amount of PDR was greater for intensity changes in an upward direction than for those in a downward direction and was greater for large intensity changes than for small intensity changes. Both of these effects were statistically significant, but there was no interaction between them. The results were interpreted to indicate that *Ss* develop preparatory sets, or expectations, regarding the stimulation they will receive and based on the stimulation they have already received. These expectations appeared to be closely tied to the physical stimulus since they were highly sensitive to changes in physical intensity.

## References

Bruner, J. S. Personality dynamics and the process of perceiving. In R. R. Blake and G. V. Ramsey, *Perception: An approach to personality*. New York: Ronald Press, 1951.

Grings, W. W., Dossett, W. F., & Honnard, R. Conditioned GSR in perceptual disparity situations. *Amer. Psychologist*, 1959, **14**, 393. (Abstract)

Grings, W. W., & Kimmel, H. D. Magnitude of GSR to expected and unexpected stimuli. Paper read at Western Psychol. Ass., Monterey, April, 1958.

Grings, W. W., & O'Donnell, D. E. Magnitude of response to compounds of discriminated stimuli. *J. exp. Psychol.*, 1956, **52**, 354–359.

Kimmel, H. D. Amount of conditioning and intensity of conditioned stimulus. Unpublished doctoral dissertation, Univer. of Southern California, 1958.

Kimmel, H. D. Amount of conditioning and intensity of conditioned stimulus. *J. exp. Psychol.*, 1959, **58**, 283–288.

(Received February 24, 1959)

**PSYCHOLOGICAL MONOGRAPHS . . . . One dollar per copy**

**No. 474. THE INFLUENCE OF DIFFERENT ACTIVITIES ON THE APPARENT LENGTH OF TIME**

By JOHN C. LOEHLIN

**No. 481. PHOTOSENSITIZATION IN RELATION TO MEAN AND STANDARD DEVIATION VALUES**

By RICHARD T. ZEGERS

---

**AMERICAN PSYCHOLOGICAL ASSOCIATION**
**1333 Sixteenth Street, N.W.**
**Washington 6, D. C.**